

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:36:43 ; Search time 53 seconds
(without alignments)
2009.195 Million cell updates/sec

Title: US-09-830-837-6
Perfect score: 5617
Sequence: 1 MKLVNWLILLVLLCGRKH.....PRVKRPQLMQVHPKPTPSV 1052

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 101223694 residues

Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5617	100.0	1052	10	US-09-060-854B-7
2	5617	100.0	1052	10	US-09-891-711-4
3	5487	97.7	1052	10	US-09-891-711-6
4	337.5	6.0	357	9	US-09-837-235-15
5	337.5	6.0	379	9	US-09-813-408-11
6	336.5	6.0	1079	9	US-10-112-488-39
7	335.5	6.0	379	10	US-09-920-118-14
8	332.5	5.9	381	10	US-09-920-118-16
9	329.5	5.9	379	9	US-09-813-408-10
10	325	5.8	380	9	US-09-813-408-16
11	324.5	5.8	379	9	US-09-813-408-13
12	323.5	5.8	275	1	US-08-322-678-8
13	323.5	5.8	275	9	US-10-033-325-2
14	323.5	5.8	275	10	US-09-060-854B-4
15	322.5	5.7	378	9	US-09-813-408-14
16	322	5.7	382	9	US-09-813-408-7
17	322	5.7	382	9	US-10-033-325-2
18	320.5	5.7	275	9	US-09-813-408-18
19	319.5	5.7	372	9	US-09-813-408-9

ALIGNMENTS

RESULT 1

US-09-060-854B-7
; Sequence 7, Application US/09060854B
; Patent No. US20020081703A1

; GENERAL INFORMATION:

; APPLICANT: Estell, David Aaron

; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical

; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002

; TITLE OF INVENTION: Proteins

; FILE REFERENCE: GC532

; CURRENT APPLICATION NUMBER: US/09/060,854B

; CURRENT FILING DATE: 1998-04-15

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 7

; LENGTH: 1052

; TYPE: PRT

; ORGANISM: Homo sapien

US-09-060-854B-7

Query Match	100.0%;	Score	5617;	DB	10;	Length	1052;		
Best Local Similarity	100.0%;	Pred. No.	0;	Mismatches	0;	Indels	0;	Gaps	0;
Matches	1052;	Conservative	0;						
Qy	1	MKLVNWLILLVLLCGKKGDLRLEKKSFEKAPCGCSHLTKVFESSTVVEYEVAF	60						
Db	1	MKLVNWLILLVLLCGKKGDLRLEKKSFEKAPCGCSHLTKVFESSTVVEYEVAF	60						
Qy	61	NGYFTAKARNSFTSSALKSSEVDNWRILIPRNNPSSDYPSPDFEVIQIKEKQAGLLTLEH	120						
Db	61	NGYFTAKARNSFTSSALKSSEVDNWRILIPRNNPSSDYPSPDFEVIQIKEKQAGLLTLEH	120						
Qy	121	PNIKRVTPQKRVPSRLKYAESDTPVPCNETRWQKWSRPLRRASLSLGSFGFWHATGRH	180						
Db	121	PNIKRVTPQKRVPSRLKYAESDTPVPCNETRWQKWSRPLRRASLSLGSFGFWHATGRH	180						
Qy	181	SSRLLRAIPRQVAQTLQADVLQMGYTGANVRVAVFDTGLSEKHPFKNVKERTNWTNE	240						
Db	181	SSRLLRAIPRQVAQTLQADVLQMGYTGANVRVAVFDTGLSEKHPFKNVKERTNWTNE	240						
Qy	241	RTLDGGLGHGTFVAGVVIASMRQCQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNALTKK	300						

Db 241 RTLDGLGHGTFVAGVIASRECOGFAPDAELHIFRFTNNQVSYTSWFLDAFNAILKK 300
Qy 301 IDVLNLSIGGPDFMDHPFVYDKWELTANNVIMVSAIGNDGPLYGTLLNPAQMDVIGVGG 360
Db 301 IDVLNLSIGGPDFMDHPFVYDKWELTANNVIMVSAIGNDGPLYGTLLNPAQMDVIGVGG 360
Qy 361 IDFDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVGGKGCALSGTSVASPVV 420
Db 361 IDFDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVGGKGCALSGTSVASPVV 420
Qy 421 AGAVTLVSTVQKRELVPASMKQALIASARLPVGNMFEQHGKLDLLRAYOILNSYKP 480
Db 421 AGAVTLVSTVQKRELVPASMKQALIASARLPVGNMFEQHGKLDLLRAYOILNSYKP 480
Qy 481 QASLSPSYDLTECPYMWPCYSQPIYYGGMTVVNTILNGMGVGRIVDKPDQWQYLPQ 540
Db 481 QASLSPSYDLTECPYMWPCYSQPIYYGGMTVVNTILNGMGVGRIVDKPDQWQYLPQ 540
Qy 541 NGDNTIEAFSYSSVLPWPSGYLAISVYTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
Db 541 NGDNTIEAFSYSSVLPWPSGYLAISVYTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
Qy 601 QTSVTKLPKVKIIPTPPRSKRVLMQDQYHNLRYPPGYPRDLNRMKNDPLDWDNGDHIHTN 660
Db 601 QTSVTKLPKVKIIPTPPRSKRVLMQDQYHNLRYPPGYPRDLNRMKNDPLDWDNGDHIHTN 660
Qy 661 FRDNTQHLRSMGYFVEVLGAPTCFDSQYGTLLMVDSEEEYFPEEIAKLRRDNDGLSL 720
Db 661 FRDNTQHLRSMGYFVEVLGAPTCFDSQYGTLLMVDSEEEYFPEEIAKLRRDNDGLSL 720
Qy 721 VIFSDWNTSVMRKRVFYDENTROWMPDGTGGANIPALNELLSSVNMGFSGLYEGETTL 780
Db 721 VIFSDWNTSVMRKRVFYDENTROWMPDGTGGANIPALNELLSSVNMGFSGLYEGETTL 780
Qy 781 ANHDMYASGCSIAKPEDGVVITOTFKDQGLEVLKQETAVVENVPILGLYQIPAEGGGR 840
Db 781 ANHDMYASGCSIAKPEDGVVITOTFKDQGLEVLKQETAVVENVPILGLYQIPAEGGGR 840
Qy 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
Db 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
Qy 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDKV 960
Db 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDKV 960
Qy 961 VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVYLAFFVQ 1020
Db 961 VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVYLAFFVQ 1020
Qy 1021 INKASRKRKRPRVKRPPOLMQOVHPKTPSV 1052
Db 1021 INKASRKRKRPRVKRPPOLMQOVHPKTPSV 1052

RESULT 2

US-09-891-711-4
; Sequence 4, Application US/09891711
; Patent No. US20020082404A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1052
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-891-711-4

Query Match Best Local Similarity 100.0%; Score 5617; DB 10; Length 1052;
Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MKLVNVLWLLVLLCGKKGHLDRLEKKSFEKAPCPGCSHLTKLVFESSTVVEYIVAF 60
Db 1 MKLVNVLWLLVLLCGKKGHLDRLEKKSFEKAPCPGCSHLTKLVFESSTVVEYIVAF 60
Qy 61 NGYFTAKARNSFTSSALKSEVDNRRIIPRNPSSDYPSPDFVIOIKKQKAGLLTLDH 120
Db 61 NGYFTAKARNSFTSSALKSEVDNRRIIPRNPSSDYPSPDFVIOIKKQKAGLLTLDH 120
Qy 121 PHIKVTPQKRVFRSLKYAESDPTVPCNETRWSQKQSSRPLRRASLSLGSQFHWAGRH 180
Db 121 PHIKVTPQKRVFRSLKYAESDPTVPCNETRWSQKQSSRPLRRASLSLGSQFHWAGRH 180
Qy 181 SSRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVFDGLSEKHPHFKNVKERTNWTNE 240
Db 181 SSRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVFDGLSEKHPHFKNVKERTNWTNE 240
Qy 241 RTLDGLGHGTFVAGVIASRECOGFAPDAELHIFRFTNNQVSYTSWFLDAFNAILKK 300
Db 241 RTLDGLGHGTFVAGVIASRECOGFAPDAELHIFRFTNNQVSYTSWFLDAFNAILKK 300
Qy 301 IDVLNLSIGGPDFMDHPFVYDKWELTANNVIMVSAIGNDGPLYGTLLNPAQMDVIGVGG 360
Db 301 IDVLNLSIGGPDFMDHPFVYDKWELTANNVIMVSAIGNDGPLYGTLLNPAQMDVIGVGG 360
Qy 361 IDFDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVGGKGCALSGTSVASPVV 420
Db 361 IDFDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVGGKGCALSGTSVASPVV 420
Qy 421 AGAVTLVSTVQKRELVPASMKQALIASARLPVGNMFEQHGKLDLLRAYOILNSYKP 480
Db 421 AGAVTLVSTVQKRELVPASMKQALIASARLPVGNMFEQHGKLDLLRAYOILNSYKP 480
Qy 481 QASLSPSYDLTECPYMWPCYSQPIYYGGMTVVNTILNGMGVGRIVDKPDQWQYLPQ 540
Db 481 QASLSPSYDLTECPYMWPCYSQPIYYGGMTVVNTILNGMGVGRIVDKPDQWQYLPQ 540
Qy 541 NGDNTIEAFSYSSVLPWPSGYLAISVYTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
Db 541 NGDNTIEAFSYSSVLPWPSGYLAISVYTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
Qy 601 QTSVTKLPKVKIIPTPPRSKRVLMQDQYHNLRYPPGYPRDLNRMKNDPLDWDNGDHIHTN 660
Db 601 QTSVTKLPKVKIIPTPPRSKRVLMQDQYHNLRYPPGYPRDLNRMKNDPLDWDNGDHIHTN 660
Qy 661 FRDNTQHLRSMGYFVEVLGAPTCFDSQYGTLLMVDSEEEYFPEEIAKLRRDNDGLSL 720
Db 661 FRDNTQHLRSMGYFVEVLGAPTCFDSQYGTLLMVDSEEEYFPEEIAKLRRDNDGLSL 720
Qy 721 VIFSDWNTSVMRKRVFYDENTROWMPDGTGGANIPALNELLSSVNMGFSGLYEGETTL 780
Db 721 VIFSDWNTSVMRKRVFYDENTROWMPDGTGGANIPALNELLSSVNMGFSGLYEGETTL 780
Qy 781 ANHDMYASGCSIAKPEDGVVITOTFKDQGLEVLKQETAVVENVPILGLYQIPAEGGGR 840
Db 781 ANHDMYASGCSIAKPEDGVVITOTFKDQGLEVLKQETAVVENVPILGLYQIPAEGGGR 840
Qy 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
Db 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
Qy 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDKV 960
Db 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDKV 960
Qy 961 VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVYLAFFVQ 1020
Db 961 VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVYLAFFVQ 1020
Qy 1021 INKASRKRKRPRVKRPPOLMQOVHPKTPSV 1052

Db 182 GSGQSWIINCIEWATSNMNOVINMSLGGTGTALKTVVDKA---VSSGIVVAAAGNE 238

QY 340 GP--LYGTLPNADQMDVIGVGGIDFEDNIARSPSRGTTWELPGYGRMKPDIVTYGAG 397

Db 239 GSGGSTVGYPAKYESTIAGVAVNSNQASFSAGS---EL-----DVMAPGV 286

QY 398 VRGSGVGGCRALSGTSVASPVVAGAVTLLVSTVQKRELVPNPKQALIASARRLPVGN 457

Db 287 IGSTLPGGTYGAVNGTSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--N 341

QY 458 MFEQGHGKLDLLRAYO 473

Db 342 SFYCKGLINVOAAAQ 357

RESULT 5

US-09-813-408-11

; Sequence 11, Application US/09813408

; Publication NO. US20030049619A1

; GENERAL INFORMATION:

; APPLICANT: Delagrave, Simon

; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides

; FILE REFERENCE: HER0041

; CURRENT APPLICATION NUMBER: US/09/813,408

; NUMBER OF SEQ ID NOS: 85

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 11

; LENGTH: 379

; TYPE: PRT

; ORGANISM: Bacillus licheniformis

US-09-813-408-11

Query Match 6.0%; Score 337.5; DB 9; Length 379;

Best Local Similarity 25.7%; Pred. No. 1.9e-20;

Matches 126; Conservative 66; Mismatches 169; Indels 129; Gaps 17;

QY 1 MKLVNILLVLLCGKKHLGRLEKSEKAPCGCSHLTLKVESSTV-----V 52

Db 2 MRKSEFWLGMTALM-----LVFTWAFSDSASAAQAKNV 36

QY 53 EYEYIVAP--NGYFTAKARNSFISSALKASVD--NRIIPNPPSSDYPDFEVIQIKKQ 110

Db 37 EKDIYVGRSGVKTASVKKDIKES--GGKVDKQFRIINAARAKLD-----80

QY 111 KAGLTLEDHPNKRVTPOKVRSLKYAESDPTVPCNETRWSQKWSRPLRRASLSLG 170

Db 81 KEALEEVNDPDV-----ATVEED-----99

QY 171 SGFWHATGRHSRRLLRAIPROVAQTLOADVLMOMGYTGANYRVAVFDVTLGSEKHPHFKN 230

Db 100 -----HYAHALAQTPYGI--PLIKADKVOAQYKGANVKVAVLDTGIQASHPDNAV 149

QY 231 VKERTNWTNERTLDDGLGHGTFVAGVIASMRQEQ---GFAPDAELHIFRVTNNQVSYTS 287

Db 150 VGSASFVAGEAYNTDGNHGTHVAGTVAALDNTTGVLGVPVNVSLYAVKVLNSSGSGSYS 209

QY 288 WFLDAFNAILKIDVLNLSIGGPDFMD--HPFVDKVKWELTANNVIMVSAIGNDGDL--Y 343

Db 210 GIVSGIEWATTNGMDVINMSLGGSPSGSTAMKQAVDNAY---ARGVVVVAAGNSGSGNT 266

QY 344 GTLNNPADOVDVIGVDIDFEDNIARSPSRGTTWELPGYGRMKPDIVTYGVRGSGV 403

Db 267 NTIGTPAKYDSVIAGVAVNSNRASFVSGA--EL-----EVMAPGAGVYSTP 314

QY 404 KGCGRALSGTSVASPVVAGAVTLLVSTVQKRELVPNPKQALIASARRLPVGNFEQCH 463

Db 315 TSYATLNGTSMASPHVAGAAALILS---RHPNLSASQVRNRLSLSTATYLG--SSFYCK 369

QY 464 GKLDLLRAYO 473

Db 370 GLINVERAAAQ 379

RESULT 6

US-10-112-488-39

; Sequence 39, Application US/10112488

; Publication NO. US20030082746A1

; GENERAL INFORMATION:

; APPLICANT: KIKUCHI, Yoshiaki

; APPLICANT: DATE, Masayo

; APPLICANT: UMEZAWA, Yukiko

; APPLICANT: YOKOYAMA, Keiichi

; APPLICANT: MATSUI, Hiroshi

; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSLUTAMINASE

; FILE REFERENCE: 219286USOCONT

; CURRENT APPLICATION NUMBER: US/10/112,488

; CURRENT FILING DATE: 2002-04-01

; PRIOR APPLICATION NUMBER: PCT/JP00/06780

; PRIOR FILING DATE: 2000-09-29

; PRIOR APPLICATION NUMBER: JP2000-280098

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: JP11-280098

; PRIOR FILING DATE: 1999-09-30

; NUMBER OF SEQ ID NOS: 70

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 39

; LENGTH: 1079

; TYPE: PRT

; ORGANISM: Streptomyces albobogriseolus

US-10-112-488-39

Query Match 6.0%; Score 336.5; DB 9; Length 1079;

Best Local Similarity 22.3%; Pred. No. 1.3e-19;

Matches 175; Conservative 113; Mismatches 284; Indels 211; Gaps 36;

QY 162 LRRASLSL-----GSGFWHA-----TGRHSRRLLRAIPROVAQTLOADV---201

Db 124 VRRTLSLNADAVTQDEAGAELEWAVTDGRTASGVAVWLDGVKASLDTSVGQIGTP 183

QY 202 -LWOMGYTGANRVAVFDTGLSEKHPHF--NVKERTNWTNERTLDDGLGHGTFVAGVIA-258

Db 184 KWEAGYDGKGVIAVLDTGVDATHPDLKGQVTSKNTSAPTTGVDVGVGHVAVIAAG 343

QY 259 ----SMRECOGAPDAELHIFRVTNNQVSYTSWFLDAFNAILKIDVLNLSIGGPDFM 314

Db 244 TGAQSKGYTKGVAPCAKILNGKVLDDAGGDSGLAGMEWAAAGADIVNLSLG--M 300

QY 315 DHPFVD---KWELTA--NNVIMVSAIGNDGLYGTLPNADQMDVIGGIDFEDNIAR 369

Db 301 DPETDPLEAAVDKLSAEKGLFAAGNEG--QSGSPGSDASALTGVAVDDKDLAD 358

QY 370 FSSRGTTWELPGYGRMKPDIVTYGVRGSGVKG-----CCRALSGTSVAS 417

Db 359 FSTGTPLR-----GDGAVKPDLTAPGVDDITAAKAGNDIAKEVGERPAGTMTISGTSMAT 413

QY 418 PVVAGAVTLLVSTVQKRELVPNPKQALIASARRLPVGNFQGHGKLDLLRAYOILNS 477

Db 414 PHVAGAAALL---EQHPPEWKAELKALGALTASTKD--QKYPFQSGRQVQVDKAITQTVI 469

QY 478 YKPOASLSPSYIDITECPYWPYPC-----SOPYYGGMPTVAVNTILNGMGVYTRIVDKP 532

Db 470 AEP-----VSLFQVQWPHADDPKPVTKLTYRNLGT--EDVTLKLTSTATG-----514

QY 533 DMQVPLPQNGDNIEVAP---SYSSVLWPMWSGYLAISISV--TKKAASWEGIAOCHVMTVA 588

Db 515 -----PKGAAPAGFTLGLASTLTVPANGTASVDVTDTRLGGAVDGTYSAYVATGA 567

QY 589 SPA-----ETESNGAEQTSVKPLPIKVIPIPPRS-----KRVL 624

Db 568 GQSVRTAAAVREVESTN-----VTLKLDKSGKATANYWAYLSGLTGLGKDRS 616

QY 625 WDOYH-----NLRYPPGYPFPR--NLRMKNDPLDNGDHHTNFRDMYQHLRSMGYFVEVL 678

FILE REFERENCE: HER0041
 ; CURRENT APPLICATION NUMBER: US/09/813.408
 ; CURRENT FILING DATE: 2001-03-21
 ; NUMBER OF SEQ ID NOS: 85
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 10
 ; LENGTH: 379
 ; TYPE: PRT
 ; ORGANISM: Bacillus licheniformis
 US-09-813-408-10

Query Match 5.9%; Score 329.5; DB 9; Length 379;
 Best Local Similarity 25.3%; Pred. No. 9.4e-20;
 Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;
 QY 1 MKLVNIWLLLVLCGKHLDRLEKSKFEKAPCGCSHLTKLVESSTV-----V 52
 DB 2 MRKKSFWLGMLTAFM-----LVFTMAFSDSASAAQPKNV 36
 QY 53 EYEYIVAF-NGYETAKARNSFTSSALKSSSEVD-NWRIIPRNPSSDYPDFEVIQKEQ 110
 DB 37 EKDIYVGFSGVTSVSKKDIKES--GGKVDKQFRILINAKAKLD-----80
 QY 111 KAGLTLEDHPNKRVTPOKRVFSLKYAESDTPVPCNETRWSQKWSRPLRRASLSLG 170
 DB 81 KEALKEVNDPDV-----AYVEED-----99
 QY 171 SGFWHATGRSSRRLLRAIPROVAOTLQADVLWQMGTGANYRVAVFDTGLSEKHPFKN 230
 DB 100 -----HYAHALAQTPYGI-PLIKADKVOAQGFGKANYKVAVLDTGQASHPDNLV 149
 QY 231 VKERTNMTNERTLDDGLGHGTFVAGVIASMRCEQ--GFAPDAELHIFRFTNNQVSYTS 287
 DB 150 VGGASVFAGEAYNTDGNHGHGTHVAGTVAALDNTTGLVGVAPSVLSYAVKVLNSSGGTYS 209
 QY 288 WFLDAFNAILKIDVNLISIGGPDFMD--HPFVDKMWELTANNVIMVSAIGNDGLP--Y 343
 DB 210 GIVSGIEWATTNGMDVINNLGPGSGSTAMKQAVDNAY---ARGVVVAAAGNSGSGNT. 266
 QY 344 GTLNNPADQMDVIGVGIDFEDNIAFSSRGMTTWELPGYGRMKPDIYTYGVRGSGV 403
 DB 267 NTIGYPKIDSVIAGVDSNRASFSSVGA--EL-----EVMAPGAGVYSTYP 314
 QY 404 KGGCRALSGTSVSPVAVAGATLLVSTVQKRELNVNPSMKAQALASARLLPGVNMFEQGH 463
 DB 315 TSYATLNGTSMASPHVAGAAALILS--KHPNLSAQVRNLSSTATYLG--SSFYTK 369
 QY 464 GKLDLLRAYQ 473
 DB 370 GLINVERAAQ 379

RESULT 10
 US-09-813-408-16
 ; Sequence 16, Application US/09813408
 ; Publication NO. US20030049619A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Delagrave, Simon
 ; APPLICANT: Marks, Barry
 ; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial Li
 ; TITLE OF INVENTION: Of Polynucleotides
 ; FILE REFERENCE: HER0041
 ; CURRENT APPLICATION NUMBER: US/09/813.408
 ; CURRENT FILING DATE: 2001-03-21
 ; NUMBER OF SEQ ID NOS: 85
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 16
 ; LENGTH: 380
 ; TYPE: PRT
 ; ORGANISM: Bacillus subtilis var. natto
 US-09-813-408-16
 Query Match 5.8%; Score 325; DB 9; Length 380;

Best Local Similarity 27.0%; Pred. No. 2.3e-19;
 Matches 119; Conservative 66; Mismatches 148; Indels 108; Gaps 20;
 QY 48 SSTVVEYEVAFNGYFTAKARNSFTSSALKSSSEVDNWRILIPRNPSSDYPDFEVIQIK 107
 DB 33 SST--EKYIVGF-----KQTSAMSSAKKKDVI-----59
 QY 108 EKOKAGLLTLE-DHPNKRVTPOKRVFSLKYAESDTPVPCNETRWSQKWSRPLRRAS 166
 DB 60 -SEKGGKVKQFKYVNAATLDEKAVKELK---KDPVA-----95
 QY 167 LSLGSGFWHATGRSSRRLLRAIPROVAOTLQADVLWQMGTGANYRVAVFDTGLSEKHP 226
 DB 96 -----YVEEDIAHEAQAQPYGICQ-IRALHSQGYTGSNKAVIDSGIDSSHP 146
 QY 227 HFKNKERTNMTNERT--LDDGLGHGTFVAGVIASMRCE--COGFAPDAELHIFRFTNN 281
 DB 147 DL-NVRGASVFPSENPYQDSSGHGTHVAGTIAALNNSIGVLGVAPASLTVAKVLDST 205
 QY 282 QVSYTSWFLDAFNAILKIDVNLISIGGPDFMD--HPFVDKMWELTANNVIMVSAIGND 339
 DB 206 GSGQYSWIINGIEWALSNNMDVINNLGPGTGSTALTATVVDKA---VSSGIYVAAAAGNE 262
 QY 340 GP--LYGTLNNPADQMDVIGVGIDFEDNIAFSSRGMTTWELPGYGRMKPDIYTYGAG 397
 DB 263 GSGSGTSTVGPYKPTSTTAVGAVNSNQASFSGVS---EL-----DVMAPGVS 310
 QY 398 VRGSGVKGCC-RALSGTSVSPVAVAGATLLVSTVQKRELNVN---ASMKAQALASARR 452
 DB 311 IQSSTPLGTYGAYNGTSMATH-VAGAAALILS-----HPTWTAQVRDLRESTATY 361
 QY 453 LFCVNMFEQGHKLDLLRAYQ 473
 DB 362 LG--NSFYKGLINVQAAQ 380

RESULT 11
 US-09-813-408-13
 ; Sequence 13, Application US/09813408
 ; Publication NO. US20030049619A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Delagrave, Simon
 ; APPLICANT: Marks, Barry
 ; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial
 ; TITLE OF INVENTION: Of Polynucleotides
 ; FILE REFERENCE: HER0041
 ; CURRENT APPLICATION NUMBER: US/09/813.408
 ; CURRENT FILING DATE: 2001-03-21
 ; NUMBER OF SEQ ID NOS: 85
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 379
 ; TYPE: PRT
 ; ORGANISM: Bacillus licheniformis
 US-09-813-408-13
 Query Match 5.8%; Score 324.5; DB 9; Length 379;
 Best Local Similarity 24.8%; Pred. No. 2.5e-19;
 Matches 123; Conservative 66; Mismatches 167; Indels 139; Gaps 17;
 QY 1 MKLVNIWLLLVLCGKHLDRLEKSKFEKAPCGCSHLTKLVESSTV-----V 52
 DB 2 MRKKSFWLGMLTAFM-----LVFTMAFSDSASAAQPKNV 36
 QY 53 EYEYIVAF-NGYETAKARNSFTSSALKSSSEVD-NWRIIPRNPSSDYPDFEVIQKEQ 110
 DB 37 EKDIYVGFSGVTSVSKKDIKES--GGKVDKQFRILINAKAKLD-----80
 QY 111 KAGLTLEDHPNKRVTPOKRVFSLKYAESDTPVPCNETRWSQKWSRPLRRASLSLG 170
 DB 81 KEALKEVNDPDV-----AYVEED-----99
 QY 171 SGFWHATGRSSRRLLRAIPROVAOTLQADVLWQMGTGANYRVAVFDTGLSEKHPFKN 230

Db 100 -----HVAHVLGQTPYGI-PLIKADKVOAGFKGANVKVAVLDTGQASHPDLNV 149
QY 231 VKERTNWTNERTLDGGLGHTFFVAGVIASMBRCQ---GFAPDAELHIFRFTNNQVSYTS 287
Db 150 VGASFVAGPAYWTDGNGHGHVAGTVAALONTGVLGVAPSVLSYAVKVLNLSGSGSYS 209
QY 288 WFLDAFNAYAILKIDVNLISIGPDF-----MDHPFVDKVMWELTANNVIMVSAIGNDG 340
Db 210 AIVSGLEWATTTGMDVINLSLGCASVSTAMKQAVDHAY-----ARGAVVSSAGNSG 261
QY 341 PL--YGTLLNPPADQMDVIGVGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGV 398
Db 262 SSGTNTIGYPAKYDSVIAVGAVDNSNRASFSSVGA---EL-----EYMAPGAGV 309
QY 399 RGSVGGCCRALSGTSSVSPVAVGAVTLVSTVQKRELNPASMKQALIASARRLPGVNM 458
Db 310 YSTYPTNTATLNGTSMASPHVAGAAALILS-----KHPNLSASQVTRLSTRTATYLG--SS 364
QY 459 FEQHGKLDLLRAYQ 473
Db 365 FSYGRGLINVEAAQA 379

RESULT 12

US-08-322-678-8
; Sequence 8, Application US/08322678
; Publication No. US20030077807A1
; GENERAL INFORMATION:
; APPLICANT: Graycar, Thomas P
; APPLICANT: Bott, Richard R
; APPLICANT: Wilson, Lori J
; TITLE OF INVENTION: Subtilisin Variants
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc
; STREET: 180 Kimball Way
; CITY: So. San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,678
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC235-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-322-678-8

Query Match 5.8%; Score 323.5; DB 1; Length 275;
Best Local Similarity 31.1%; Pred. No. 1.8e-19;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;
QY 187 RAIPROVATQADVLQWQGYTGANVRVAVDTGLSEKHPKFNKERTNWTNERT--LD 244
Db 2 QSVPYGISQ-IPKALHSQGYTGSNKNVAVDSGIDSSHPDL-NVRGGASFVPSNTPYQ 59

QY 245 DGLGHGTFVAGVIASMBRE---COGFAPDAELHIFRFTNNQVSYTSWFLDAFNAYAILKKI 301
Db 60 DGSSHGTHVAGTIAALNNSIGVLSVSPASLYAVKVLDTGSGQYSWIINGIEWALSNNM 119
QY 302 DVNLNLSIGGPDFMD--HPFVDKVMWELTANNVIMVSAIGNDGP--LYGTLLNPPADQMDVIG 357
Db 120 DVINMSLGGTGTSTALKTVVDKA---VSSGIVVAAAAGNEGSSGSTSTVGYPAKYPSTIA 176
QY 358 VGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVKGCGCRALSGTSVAS 417
Db 177 VGVANSSNRASFSSAGS---EL-----DVMAPGVSIQSTLPGGTGYGAYNGTSMAT 224
QY 418 PVVAGAVTLVSTVQKRELNPASMKQALIASARRLPGVNMFGQHGKLDLLRAYQ 473
Db 225 PHVAGAAALILS---KHPWTNAQVRDLRLESTATYLG--NSFYFGKGLINVOAAAQ 275

RESULT 13

US-10-033-325-4
; Sequence 4, Application US/10033325
; Publication No. US20030073222A1
; GENERAL INFORMATION:
; APPLICANT: Poulos, Ayrookaran J.
; APPLICANT: Schellenberger, Volker
; APPLICANT: Kellis, Jr., James T.
; APPLICANT: Paech, Christian
; APPLICANT: Nadherny, Joanne
; APPLICANT: Naki, Donald P.
; APPLICANT: Collier, Katherine D.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Baeck, Andre C.
; TITLE OF INVENTION: Multiply-Substituted Protease Variants
; FILE REFERENCE: GC502-2-C1
; CURRENT APPLICATION NUMBER: US/10/033,325
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: US 09/178,155
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: US 08/956,323
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,564
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,324
; PRIOR FILING DATE: 1997-10-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-033-325-4

Query Match 5.8%; Score 323.5; DB 9; Length 275;
Best Local Similarity 31.1%; Pred. No. 1.8e-19;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;
QY 187 RAIPROVATQADVLQWQGYTGANVRVAVDTGLSEKHPKFNKERTNWTNERT--LD 244
Db 2 QSVPYGISQ-IPKALHSQGYTGSNKNVAVDSGIDSSHPDL-NVRGGASFVPSNTPYQ 59
QY 245 DGLGHGTFVAGVIASMBRE---COGFAPDAELHIFRFTNNQVSYTSWFLDAFNAYAILKKI 301
Db 60 DGSSHGTHVAGTIAALNNSIGVLSVSPASLYAVKVLDTGSGQYSWIINGIEWALSNNM 119
QY 302 DVNLNLSIGGPDFMD--HPFVDKVMWELTANNVIMVSAIGNDGP--LYGTLLNPPADQMDVIG 357
Db 120 DVINMSLGGTGTSTALKTVVDKA---VSSGIVVAAAAGNEGSSGSTSTVGYPAKYPSTIA 176
QY 358 VGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVKGCGCRALSGTSVAS 417
Db 177 VGVANSSNRASFSSAGS---EL-----DVMAPGVSIQSTLPGGTGYGAYNGTSMAT 224
QY 418 PVVAGAVTLVSTVQKRELNPASMKQALIASARRLPGVNMFGQHGKLDLLRAYQ 473

Qy	258	ASMBRE---	CCGFAPDAELHIFRVTNNNOVSYTSWFELDAFNAYILAKKIDVNLNLSIGGPDFM	314
Db	184	AALNSIGVGVGAPDAEYIAVKVLGANGSGSVSSIAOGLWQAQNNIHVANSLSGS---	239	
Qy	315	DHPFVDKWEELTAN----	NVMKYSIAIGNDGPLYGTLLNPDQMDMDVIGGGIDFPEDNTAR	369
Db	240	--PVGSOTLELAVNQARNAGVLLVVAATGNGSG--	GTVSYPARYANALAVGATDQNNRRAS	295
Qy	370	FSSRGMTTWELPGGYGRMKPDIIVTYGAGVSGVKGCCRALSGTSSVSPVVGAVTLVLS	429	
Db	296	FSQYGTGL-----	NIVAPGVIQSTYPCNRYASLSGTSMATPHVAGVAAVLV---	341
Qy	430	TVQKRELVPNSMKQALIASARRLPFGVMFQQG	462	
Db	342	--KOKNPWSNFOIRHOHLTSTATSLGNSNOFGSG	373	

RESULT 17

```

US-10-033-325-2
; Sequence 2, Application US/10033325
; Publication No. US2003007322A1
; GENERAL INFORMATION:
; APPLICANT: Poulouze, Avrookaran J.
; APPLICANT: Schellenberger, Volker
; APPLICANT: Kellis, Jr., James T.
; APPLICANT: Pasch, Christian
; APPLICANT: Nadherny, Joanne
; APPLICANT: Naki, Donald P.
; APPLICANT: Collier, Katherine D.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Baeck, Andre C.
; TITLE OF INVENTION: Multiply-Substituted
; FILE REFERENCE: G502-2-cl
; CURRENT APPLICATION NUMBER: US/10/033,325
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: US 09/178,155
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: US 08/956,323
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,564
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,324
; PRIOR FILING DATE: 1997-10-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-10-033-325-2

```

Query Match	5.7%	Score 322;	DB 9;	Length 382;
Best Local Similarity	26.2%;	Pred. No. 4.2e-19;		
Matches 119; Conservative	57;	Mismatches 142;	Indels 136;	Gaps 18;
QY	1	MLKLVNWLILLVLLCGKKHGLDRLEKKSEFAPCPGCSHLTLKVPESSIVV-----	52	
Db	1	MRGKKVMSILFLA-----LITMAGFSTSSAQAAQAKSN 35		
QY	53	-EYEIVAFNGYFTAKARNISFISSALKSSVDNWRIIPRNPSSDPDSFEVIQIKEKQ 111		
Db	36	GEKKYIVGF-----KQTWSTMSAAKKDVI-----SEK 63		
QY	112	AGLLTLE-DHPNIKRVTPOQRVPSRLKYAESDTPVCNETRWQSKWQSRRPLRRASLSLG 170		
Db	64	GGKVQKFQYVDAASTVLTNEKAVKELK-----KDPSVA-----	96	
QY	171	SGFWHATGRHSRLLLRATPROVAQTLQADVLWMQGYTGANRVAVFDTCGLSEKHPHFK- 229		
Db	97	----YYEEDHVAHAYQAQSPYPGVSQ-IKAPALHSQGTYGTGSNVKAVVIDSIDSSHDLKV 151		
QY	230	----NVKEKTNTNERTLDDGLGHGHTFFVAGVIASMRE-----COGFAPDAELHIHFVFPTNN 281		

152 ASGASMPSETN-----PFQDNNSHGTHVAGTVAAALNNSIGVLGVAPSASLVAVKVLGAD 206

282 QVSYTSWFLDAFNVAIILKKIDVLNLSIGGPDFMD--HPFVDKVWELTANNVIMVSAGIND 339

207 GSGOYSWINGIEWALANNMDVLNMSLGGPSCSAALKAAVDKA---VSGVVVVVAAGNE 263

340 GP--LYGTNLNPADQMDVTVGVGIDEDFNIAFSSRGMTTWELPGCYGRMKKPDITYYGAG 397

264 GTSGSSSTVGYPGPKYPSVIAVGAVDSSNORASFSSVGP---EL-----DVMAPGVS 311

398 VRGS--GVKGGCRALSGTSVASPVVAGAVTLLVS 429

312 IQSTLPGNKYG--AYNGTSMASPHVAGAAAILLS 343

RESULT 18

US-09-813-408-18

; Sequence 18, Application US/09813408

; Publication No. US20030049619A1

RESULT 18

```

US-09-813-408-18
; Sequence 18, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maris, Barry
; TITLE OF INVENTION: Methods For The Synthesis
; OF Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus pumilus
US-09-813-408-18

```

```

Query Match      5.7%; Score 320.5; DB 9; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.3e-19;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

187 RAIPROVAOTLQADVLWQMGCYANRVAVFDTGLSEKHPKFNKERTNWTNERT--LD 244
   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2 QSVPYGISQ-IRAPALHSQGYTGSNKKVAVIDSGIDSSHPDL-NVRGGASVPSETNPQY 59

245 DGLGHGCTFVAGVYATSMRE---COGFAPDAELHFRVFTNNQSYYSWFLDAFNVALKKI 301
   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
60 DGSSGTHVAGTITRALNNSIGVLGVAPSASLFAVKVLDSTGSGQYSWIINGTAWLSNNM 119

302 DVNLNLSIGGPDFMD--HPFVDKVKWELTANNVIMVSAIGNDGP--LYGTLLNPADQMDVIG 357
   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
120 DVINMSLGGPTGSLTKTVVDKA---VSSGIVVAAAAGNEGSSGSTSTVGYPAPKYSTIA 176

358 VGSIGDPEDNIAFFSSRGMTTWELPGCYGRMKPDIITYGAGVRGSGVKGCRLSGTSVAS 417
   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
177 VGANVSANQASFSNAGS---EL-----DVWAPGVSIQSTLPGGTGYGAYNGTSMAT 224

418 PVVAGAVTLLVSTVQKRELNVPSMKQALIIASARRLPGVNMFEGQHGKLDLRLAYQ 473
   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
225 PHVGAAGAAIILS---KHPTWTWNAQVRDLRESTATYLG--SSFYQKGLINVQAAAQ 2775
   :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

```

RESULT 19

```

RESOLUTION 13
US-09-813-408-9
; Sequence 9, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maris, Barry
; TITLE OF INVENTION: Methods For The Synthesis
; TITLE OF INVENTION: Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-813-408-9

Query Match 5.7%; Score 319.5; DB 9; Length 372;
Best Local Similarity 24.6%; Pred. No. 6.6e-19;
Matches 119; Conservative 66; Mismatches 170; Indels 129; Gaps 16;
QY 1 MKLVNIWLLLVLLGCKHGLDRLEKKSPKAPCGCSHITLKVFEFSVTV-----V 52
DB 2 MRKKSFLGMLTAFM-----LVETMAFSDSASAAQPAKNV 36
QY 53 EYEVITAF-NGYFTAKARNSFISSALKSSEVDNWRRIIPRNPSSDYPDFEVIQIKQK 111
DB 37 EKDIYVFKSGVKTASVKDKIIES--GGKVD-----KQPRIIN-----73
QY 112 AGLLTLEDHPIKIRVTPQKVFERSIKYAESDPVPCNETRWSQKWSSRLRASLSLGS 171
DB 74 -----AAKAKIDEALKEVKNDPVAYEED-----97
QY 172 GWHATGHSRRLRAIPROVAOTLOADVLMQGYTGANVRVAVFTGLSEKHPHFKNV 231
DB 98 -----HVAHALAQTVPIGL-PLIRADKVOAQGFGKANVKVAVLDTGQASHPDLNV 148
QY 232 KERTNWTNERTLDGLGHGTFVAGVIASMRQ---GFAPDAELHIFRFTNNQVSYTSW 288
DB 149 GGASEVAGEAYNTDNGHGHGTHVAGTVAALDNTTGVLGAVPSVLYAVKVLNLSGSGSYSG 208
QY 289 FIDAFNYAILKKIDVLNLSIGGPDFMD--HPFVDKVMWELTANNVIMVSAIGNDGL--YG 344
DB 209 IVSGIEWATNGMDVINNSLGGASGSGTAMQAVDNAY---AKGVVVVAAAGNSGSGNTN 265
QY 345 TLNNPADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVK 404
DB 266 TIGYPKDYSVIAGVADSNRASFSSVGA---EL-----EVMAPGAGVSYTPT 313
QY 405 GCRALSGTSVASPVVAGAVTLLVSTVKRELVPASMKQALIASARRLPVGNMFEQHG 464
DB 314 NYATLNGTSMASPHVAGAAALILS---KHPNLTASQVRNLSSTATYLG--SSFYKKG 368
QY 465 KLDL 468
DB 369 LINV 372

RESULT 20
US-10-090-624-31
; Sequence 31, Application US/10090624
; Patent No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090.624
; PRIOR FILING DATE: 2002-03-06
; PRIOR FILING DATE: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-090-624-31

Query Match 5.7%; Score 319.5; DB 12; Length 382;
Best Local Similarity 31.7%; Pred. No. 6.9e-19;
Matches 98; Conservative 50; Mismatches 118; Indels 43; Gaps 13;
QY 180 HSSRLLRAIPROVAOTLOADVLMQGYTGANVRVAVFTGLSEKHPHFK-----NVKE 233
DB 102 HVAHAYAQSPVPGVQ-IPALHSGQYTGNSNVKVAVIDSGIDSSHPLDKVAGGASWVPS 160
QY 234 RTNWTNERTLDGLGHGTFVAGVIASMR-----COGFAPDAELHIFRFTNNQVSYTSWFL 290
DB 161 ETN-----PFQDNNSHGTHVAGTVAALNNSIGVLGVAPSASLYAVKVLGADGSGQYSWII 215
QY 291 DAFNYAILKKIDVLNLSIGGPDFMD--HPFVDKVMWELTANNVIMVSAIGNDGP--LYGTL 346
DB 216 NGIEWATANNMDVINNSLGGPSAALKAADKA---VASGVVVVAAAGNEGTSGSSTV 272
QY 347 NNPADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRG--CVK 404
DB 273 GYPGKPSYIAGVADSSNQASFSVGP---EL-----DVMAPGVSIQSTLPGNK 320
QY 405 GCRALSGTSVASPVVAGAVTLLVSTVKRELVPASMKQALIASARRLPVGNMFEQHG 464
DB 321 YG--AYNGTSMASPHVAGAAALILS---KHPNWTNTOVRSLENTTTLG--DSFYKKG 373
QY 465 KLDLRAVQ 473
DB 374 LINVQAAQ 382

RESULT 21
US-09-813-408-19
; Sequence 19, Application US/09813408
; Publication NO. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maers, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-09-813-408-19

Query Match 5.7%; Score 319; DB 9; Length 380;
Best Local Similarity 31.3%; Pred. No. 7.6e-19;
Matches 96; Conservative 52; Mismatches 119; Indels 40; Gaps 12;
QY 180 HSSRLLRAIPROVAOTLOADVLMQGYTGANVRVAVFTGLSEKHPHFK-----NVKE 233
DB 101 HVAHAYAQSPVPGVQ-IPALHSGQYTGNSNVKVAVIDSGIDSSHPLDKVAGGASWVPS 159
QY 234 RTNWTNERTLDGLGHGTFVAGVIASMR-----COGFAPDAELHIFRFTNNQVSYTSWFL 290
DB 160 ETN-----PFQDNNSHGTHVAGTVAALNNSIGVLGVAPSASLYAVKVLGADGSGQYSWII 214
QY 291 DAFNYAILKKIDVLNLSIGGPDFMDHPFVDKVMWELTANNVIMVSAIGNDGP--LYGTLN 348
DB 215 NGIEWATANNMDVINNSLGGPS--GSAALKAVDKAVASGVVVVAAAGNEGTSGSSTVGY 272
QY 349 PADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRG--GVKGG 406
DB 273 PGKPSYIAGVADSSNQASFSVGP---EL-----DVMAPGVSIQSTLPGNKY 320
QY 407 CEALSGTSVASPVVAGAVTLLVSTVKRELVPASMKQALIASARRLPVGNMFEQHGKL 466
DB 321 --AYNGTSMASPHVAGAAALILS---KHPNWTNTOVRSLENTTTLG--DSFYKGLI 373

QY 467 DLLRAYQ 473
 Db 374 NVQAAQ 380

RESULT 22
 US-09-837-235-17
 ; Sequence 17, Application US/09813408
 ; Patent No. US20030049619A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Marshall, Christopher
 ; APPLICANT: Hoffman, Alexander
 ; APPLICANT: Errico, Joseph
 ; APPLICANT: Marshall, Paul
 ; TITLE OF INVENTION: STABILIZED PROTEINS
 ; FILE REFERENCE: 9725-005-999
 ; CURRENT APPLICATION NUMBER: US/09/837,235
 ; CURRENT FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: PCT/US00/28595
 ; PRIOR FILING DATE: 2000-10-16
 ; PRIOR APPLICATION NUMBER: 60/159,763
 ; PRIOR FILING DATE: 1999-10-15
 ; NUMBER OF SEQ ID NOS: 42
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 17
 ; LENGTH: 275
 ; TYPE: PRT
 ; ORGANISM: Bacillus subtilis
 US-09-837-235-17

Query Match 5.7%; Score 318.5; DB 10; Length 275;
 Best Local Similarity 30.7%; Pred. No. 5e-19;
 Matches 91; Conservative 55; Mismatches 119; Indels 31; Gaps 11;

QY 187 RAIPROVAQTQADVLQMGYTGANVRVAVFTDGLSEKHPKFNKERTNWTNERT--LD 244
 Db 2 QSVPYGISQ-IKAPALHSQGYTGSNVKVAVIDSGIDSSHPDL-NVRGGASFVPSVETNPQ 59
 QY 245 DGLGHTFTVAGVIASMR--HPFVDKRWELTANNVIMVSAIGNDGP--LYGTLLNPPADQMDVIG 357
 Db 60 DVNNSLGGPTGTALKTVVDKA---VSSGIIVAAAGNEGSGSTSTVGYPAKYPSTIA 176
 QY 358 VGGIDFEDNIARESSRGMTTWELPGGYGRMKPDIVYAGVRSYVGGCRALSGTSVAS 417
 Db 177 VGAVNSSNQRASFSSAGS---EL-----DVMAPGVSIQSTLPGGYGAYNGTCMAT 224
 QY 418 PVVAGAVTLLVSTVQKRELNVNPMQKALIASARLPGVNMFEQGHGKLDLLRAYQ 473
 Db 225 PHVGAALILS---KHPTWNAQVRDRLESTATYLG--NSFYGKGLINVOAAQ 275

RESULT 23
 US-09-813-408-17
 ; Sequence 17, Application US/09813408
 ; Publication No. US20030049619A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Delagrave, Simon
 ; APPLICANT: Maits, Barry
 ; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial Li
 ; TITLE OF INVENTION: Of Polynucleotides
 ; FILE REFERENCE: HER0041
 ; CURRENT APPLICATION NUMBER: US/09/813,408
 ; CURRENT FILING DATE: 2001-03-21
 ; NUMBER OF SEQ ID NOS: 85
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 17
 ; LENGTH: 274
 ; TYPE: PRT

Query Match 5.6%; Score 317; DB 9; Length 274;
 Best Local Similarity 30.8%; Pred. No. 6.6e-19;
 Matches 91; Conservative 54; Mismatches 120; Indels 30; Gaps 1

QY 187 RAIPROVAQTQADVLQMGYTGANVRVAVFTDGLSEKHPKFNKERTNWTNERT--LD 244
 Db 2 QSVPYGISQ-IKAPALHSQGYTGSNVKVAVIDSGIDSSHPDL-NVRGGASFVPSVETNPQ 59
 QY 245 DGLGHTFTVAGVIASMR--HPFVDKRWELTANNVIMVSAIGNDGP--LYGTLLNPPADQMDVIG 357
 Db 60 DVNNSLGGPTGTALKTVVDKA---VSSGIIVAAAGNEGSGSTSTVGYPAKYPSTIA 176
 QY 358 VGGIDFEDNIARESSRGMTTWELPGGYGRMKPDIVYAGVRSYVGGCRALSGTSVAS 417
 Db 177 VGAVNSSNQRASFSSAGS---EL-----DVMAPGVSIQSTLPGGYGAYNGTCMAT 224
 QY 419 VVAGAVTLLVSTVQKRELNVNPMQKALIASARLPGVNMFEQGHGKLDLLRAYQ 473
 Db 225 HVGAAALILS---KHPTWNAQVRDRLESTATYLG--NSFYGKGLINVOAAQ 274

RESULT 24
 US-10-090-624-12
 ; Sequence 12, Application US/10090624
 ; Patent No. US20020132335A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: SHIMOJO, Tomoko
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 ; FILE REFERENCE: TAKAKURA-6
 ; CURRENT APPLICATION NUMBER: US/10/090,624
 ; CURRENT FILING DATE: 2002-03-06
 ; PRIOR APPLICATION NUMBER: 09/445,472
 ; PRIOR FILING DATE: 1999-12-06
 ; PRIOR APPLICATION NUMBER: 151969/1997
 ; PRIOR FILING DATE: 1997-06-10
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 659
 ; TYPE: PRT
 ; ORGANISM: Thermococcus celer
 US-10-090-624-12

Query Match 5.6%; Score 316; DB 12; Length 659;
 Best Local Similarity 26.8%; Pred. No. 3.3e-18;
 Matches 126; Conservative 65; Mismatches 192; Indels 88; Gaps 2

QY 199 ADVLW-QMGYTGANVRVAVFTDGLSEKHPKFNKERTNWTN-----TNERFLDGLGHGTFV 25
 Db 147 ADTVNNSLGGYDGSVVAIVDTGIDANHPDLKG--KVIGWYDAVNGRSTPYDDQGHGTHV 20
 QY 254 AGVIASMR--HPFVDKRWELTANNVIMVSAIGNDGP--LYGTLLNPPADQMDVIG 357
 Db 205 AGIVAGTGSVNSQYIGVAPGAKLVGVKLGADSGSVSTIIAGVDMVQVQNKDKYGIK 26
 QY 306 LSGISG-----GPDFMDHPFVKRWELTANNVIMVSAIGNDGP--LYGTLLNPPADQMDVIG 357
 Db 265 LSLGSSQSDGTDLSLQA-VNNAWDA---GIVCVVAGNSGNTTVGSPRAASKVIIVG 32
 QY 360 GIDFEDNIARESSRGMTTWELPGGYGRMKPDIVYAGV---RGSYKGGC-----RAL 41
 Db 321 AVDSNDNIASFSSRG-----PTADGRKLPVAVPGVDIIAPRASGTSNGTPIINDYTTKA 37

116 ANNDVFNMSLGPSSAALAAVDKA---VAGSVVVAAGNEGTSSTVGPYKYP 172
354 DVICVGVGIDFEDNIARFSSRGMTTWELPGYGRMKPDIVTYCAGYRGS--GVKGGCRALS 411
173 SVIAGVAVDSSNQASFSVGP---EL-----DVMAPGVISQSTLPGNKYG--AYN 218
412 GTSVASPVVAGAVTLLVSTVOKRELVPASMKQALIASARRLPVGNMPEQCHGKLDLLRA 471
219 GTSNASPHVAGAAALILS---KHPNWTNTQVRSLENTTKLG---DSFYKGLINVOAA 273
472 YQ 473
274 AQ 275
Search completed: May 29, 2003, 13:41:25
Job time : 56 secs

411 SGTSPVAVAGAVTLLVSTVOKRELVPASMKQALIASARRLPVGNMPE--QHGKLDL 468
375 SGTSMATPHVSGVALI---LQAHPSWTDPKVKTALETADIAPKEIADIAGAGRVN 431
469 LRAVQILNSYKPOASL--SPSYIDLTECPYMPYCQPIYYGGMPTVVNVTTILNGMVTG 526
432 YKAKI---YDDYAKLTFTGSAVDKGSATHTDVSQ-----AUFVTATLYWDTG--- 476
527 RIVKDPQWQYL-PONGNIEVAES-----YSSVLPWPSGYLAISISVTKKAASW 575
477-----SSDILYLDPNNGNEVDYSYATYGFYKGVYNFTAGTW-----TVKVVSY 522
576 EGIAGQHVMTVAPAEIESKNGAEQSTVKLPKIKVLIPTPPRSKRYLWD 626
523 KGAANYQDV-VSDGSLSGSGGNPNPNPNTPTTDTQTFTGTVNDYWD 572

RESULT 25
US-08-322-678-7
; Sequence 7, Application US/08322678
; Publication No. US20030077807A1
; GENERAL INFORMATION:
; APPLICANT: Graycar, Thomas P
; APPLICANT: Bott, Richard R
; APPLICANT: Wilson, Lori J
; TITLE OF INVENTION: Subtilisin Variants
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc
; STREET: 180 Kimball Way
; CITY: So. San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,678
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC235-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-322-678-7

Query Match 5.6%; Score 314.5; DB 1; Length 275;
Best Local Similarity 32.1%; Pred. No. 1.le-18;
Matches 97; Conservative 49; Mismatches 113; Indels 43; Gaps 13;
QY 187 RAIPROVACTLOADVLAQMGYTGANVRVAVFDGLSEKHPFK-----NVKERTNWTNE 240
Db 2 QSPVYGSQ-IKAPALHSQGYTGSNVKVAVIDSGIDSHDPLKAVAGASWYVSETN---- 56
QY 241 RTLDDGLGHGFVAGVIASMRE---COGFAPDAELHIFVFTNNQVSYTWSFLDAFNAL 297
Db 57 -PFQDNNSHGTHVAGTVAALNNSIGVLGVAFPSLAVKVLGADSGSGYSWINGIEWAI 115
QY 298 LKKIDVLNLSIGGDFEMD--HPFVDKVMELTANNVIMVSAIGNDGP--LYGTLLNPDQM 353

Mon Jun 2 11:46:23 2003

us-09-830-837-6.ra1

Page 2

QY 361 IDFEEDNIAFSSRGMTTWELPGYGRMKPDIVYGAAGVSGVKGCRALSGTSVASPVY 420
DB 361 IDFEEDNIAFSSRGMTTWELPGYGRMKPDIVYGAAGVSGVKGCRALSGTSVASPVY 420
QY 421 AGAVTLIVSTVOKRELVPASMKOALIASARLPGVMMFPOGHGKLDLIRAYOILNSTR 480
DB 421 AGAVTLIVSTVOKRELVPASMKOALIASARLPGVMMFPOGHGKLDLIRAYOILNSTR 480
QY 481 QASISPSYIDLTECPYMPYCOPITYYGMPVTVNVTILNGMVGRIYDNDPMDQPLPQ 540
DB 481 QASISPSYIDLTECPYMPYCOPITYYGMPVTVNVTILNGMVGRIYDNDPMDQPLPQ 540
QY 541 NGDNIEVAFSSYVLMWMSGYLAISVTKKAASMEGIAOGHMTIVASPAETESKNGAE 600
DB 541 NGDNIEVAFSSYVLMWMSGYLAISVTKKAASMEGIAOGHMTIVASPAETESKNGAE 600
QY 601 QSTVVKLPITVKIIPPPRSKRYLMDQYHNLRYPPGYFPBDNLRMKNDPLMNGDHIHTN 660
DB 601 QSTVVKLPITVKIIPPPRSKRYLMDQYHNLRYPPGYFPBDNLRMKNDPLMNGDHIHTN 660
QY 661 FPDWYOHLSRSMGYFVEVLGAFPCFASOYGTLLMVDSEEEYEPERIALRRDYDNGLSL 720
DB 661 FPDWYOHLSRSMGYFVEVLGAFPCFASOYGTLLMVDSEEEYEPERIALRRDYDNGLSL 720
QY 721 VIFSDWYNTSVARKRYFYDENTROMMPTDGGANI PALNELLSVNMGFSOGLYEGEFTL 780
DB 721 VIFSDWYNTSVARKRYFYDENTROMMPTDGGANI PALNELLSVNMGFSOGLYEGEFTL 780
QY 781 ANHDMYASGCSIAKPEDGVYITQFPOGLYVKOETAVENVYILGLYOIPAEGBGR 840
DB 781 ANHDMYASGCSIAKPEDGVYITQFPOGLYVKOETAVENVYILGLYOIPAEGBGR 840
QY 841 IYVYGDSCNCLDSSHROKCFWLMDALLOYSYGVTPPSLSHNSNRORPPSGAGSVTERM 900
DB 841 IYVYGDSCNCLDSSHROKCFWLMDALLOYSYGVTPPSLSHNSNRORPPSGAGSVTERM 900
QY 901 EGNHILHRYSKYLEAHLDGPKRPLPACPRLSNAKPOPLNETAPSNLMKHOKLLSIDDKV 960
DB 901 EGNHILHRYSKYLEAHLDGPKRPLPACPRLSNAKPOPLNETAPSNLMKHOKLLSIDDKV 960
QY 961 VLPNFRSNRPQVRPLSPGESGAMIDPGIMPGRYNOEVQOTIVFAFLGAMVYLAFFVYQ 1020
DB 961 VLPNFRSNRPQVRPLSPGESGAMIDPGIMPGRYNOEVQOTIVFAFLGAMVYLAFFVYQ 1020
QY 1021 INKASRRPRRRKPRVKRPOQLMQOVHPKTPSV 1052
DB 1021 INKASRRPRRRKPRVKRPOQLMQOVHPKTPSV 1052

RESULT 2
US-09-360-237-3
Sequence 3 Application US/09360237
Patent No. 632962

GENERAL INFORMATION
APPLICANT: BROWN, MICHAEL S.
APPLICANT: CHENG, DONG
APPLICANT: ESPENSHADE, PETER J.
APPLICANT: GOLDSTEIN, JOSEPH L.
APPLICANT: RAMSON, ROBERT B.
APPLICANT: SAKAI, JURO
TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF
FILE REFERENCE: UTXD:567
CURRENT APPLICATION NUMBER: US/09/360,237
EARLIER FILING DATE: 1999-07-23
EARLIER APPLICATION NUMBER: 60/096,571
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1052
TYPE: PRT

ORGANISM: Human
US-09-360-237-3
Query Match 100.0%; Score 5617; DB 4; Length 1052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKVNTMLLVLLVLCCKRHLGDRLEKSPKAPCGCSHLLTKYESSTVYEXYVAF 60
DB 1 MKVNTMLLVLLVLCCKRHLGDRLEKSPKAPCGCSHLLTKYESSTVYEXYVAF 60
QY 61 NGFTAKANSEFISALKSEEDVNRKIIIPRNNPSDDPSDFEYIQKQKAGLILDEH 120
DB 61 NGFTAKANSEFISALKSEEDVNRKIIIPRNNPSDDPSDFEYIQKQKAGLILDEH 120
QY 121 PNIRKVTPOKRVFSLKYASDPTVPCNETRMSOKWOSRPLRASLSLSGFWHATGNH 180
DB 121 PNIRKVTPOKRVFSLKYASDPTVPCNETRMSOKWOSRPLRASLSLSGFWHATGNH 180
QY 181 SSRRLLRAIPROVADLOADVIMQMGYGANVRAVDFGLSEKHPRKVKERTWINE 240
DB 181 SSRRLLRAIPROVADLOADVIMQMGYGANVRAVDFGLSEKHPRKVKERTWINE 240
QY 241 RTLDDELGHGTFVAGVIAISMRECOGFAPDAELHIFRVFTNNOVSTYSWFLDAFNVAILK 300
DB 241 RTLDDELGHGTFVAGVIAISMRECOGFAPDAELHIFRVFTNNOVSTYSWFLDAFNVAILK 300
QY 301 IDVLNLSIGGPFDMHDFVDKVELFANNVIMVSAIGNDPLYGLTLNRPADQMDVIGVG 360
DB 301 IDVLNLSIGGPFDMHDFVDKVELFANNVIMVSAIGNDPLYGLTLNRPADQMDVIGVG 360
QY 361 IDPEDNIAFSSRGMTTWELPGYGRMKPDIVYGAAGVSGVKGCRALSGTSVASPVY 420
DB 361 IDPEDNIAFSSRGMTTWELPGYGRMKPDIVYGAAGVSGVKGCRALSGTSVASPVY 420
QY 421 AGAVTLIVSTVOKRELVPASMKOALIASARLPGVMMFPOGHGKLDLIRAYOILNSTR 480
DB 421 AGAVTLIVSTVOKRELVPASMKOALIASARLPGVMMFPOGHGKLDLIRAYOILNSTR 480
QY 481 QASISPSYIDLTECPYMPYCOPITYYGMPVTVNVTILNGMVGRIYDNDPMDQPLPQ 540
DB 481 QASISPSYIDLTECPYMPYCOPITYYGMPVTVNVTILNGMVGRIYDNDPMDQPLPQ 540
QY 541 NGDNIEVAFSSYVLMWMSGYLAISVTKKAASMEGIAOGHMTIVASPAETESKNGAE 600
DB 541 NGDNIEVAFSSYVLMWMSGYLAISVTKKAASMEGIAOGHMTIVASPAETESKNGAE 600
QY 601 QSTVVKLPITVKIIPPPRSKRYLMDQYHNLRYPPGYFPBDNLRMKNDPLMNGDHIHTN 660
DB 601 QSTVVKLPITVKIIPPPRSKRYLMDQYHNLRYPPGYFPBDNLRMKNDPLMNGDHIHTN 660
QY 661 FPDWYOHLSRSMGYFVEVLGAFPCFASOYGTLLMVDSEEEYEPERIALRRDYDNGLSL 720
DB 661 FPDWYOHLSRSMGYFVEVLGAFPCFASOYGTLLMVDSEEEYEPERIALRRDYDNGLSL 720
QY 721 VIFSDWYNTSVARKRYFYDENTROMMPTDGGANI PALNELLSVNMGFSOGLYEGEFTL 780
DB 721 VIFSDWYNTSVARKRYFYDENTROMMPTDGGANI PALNELLSVNMGFSOGLYEGEFTL 780
QY 781 ANHDMYASGCSIAKPEDGVYITQFPOGLYVKOETAVENVYILGLYOIPAEGBGR 840
DB 781 ANHDMYASGCSIAKPEDGVYITQFPOGLYVKOETAVENVYILGLYOIPAEGBGR 840
QY 841 IYVYGDSCNCLDSSHROKCFWLMDALLOYSYGVTPPSLSHNSNRORPPSGAGSVTERM 900
DB 841 IYVYGDSCNCLDSSHROKCFWLMDALLOYSYGVTPPSLSHNSNRORPPSGAGSVTERM 900
QY 901 EGNHILHRYSKYLEAHLDGPKRPLPACPRLSNAKPOPLNETAPSNLMKHOKLLSIDDKV 960
DB 901 EGNHILHRYSKYLEAHLDGPKRPLPACPRLSNAKPOPLNETAPSNLMKHOKLLSIDDKV 960
QY 961 VLPNFRSNRPQVRPLSPGESGAMIDPGIMPGRYNOEVQOTIVFAFLGAMVYLAFFVYQ 1020
DB 961 VLPNFRSNRPQVRPLSPGESGAMIDPGIMPGRYNOEVQOTIVFAFLGAMVYLAFFVYQ 1020

Db 961 VLPNFRSNRPQVRLSPGSGAMDPGIMGRYNOEVGOTIPVAFAGAVYLAFFVQ 1020
QY 1021 INKAKSRKRRKPRVKRPOLMOQVHPKPTSV 1052
Db 1021 INKAKSRKRRKPRVKRPOLMOQVHPKPTSV 1052

RESULT 3

US-09-360-237-1
Sequence 1, Application US/09360237
Patent No. 6322962
GENERAL INFORMATION:
APPLICANT: BROWN, MICHAEL S.
APPLICANT: CHENG, DONG
APPLICANT: ESPENSHADE, PETER J.
APPLICANT: GOLDSTEIN, JOSEPH L.
APPLICANT: RAMSON, ROBERT B.
APPLICANT: SAKAI, JURO
TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF
TITLE OF INVENTION: MODULATORS THEREOF
FILE REFERENCE: 07XD:567
CURRENT APPLICATION NUMBER: US/09/360, 237
CURRENT FILING DATE: 1999-07-23
EARLIER APPLICATION NUMBER: 60/096, 571
EARLIER FILING DATE: 1998-08-14
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1052
TYPE: PRT
ORGANISM: Cricetus griseus
US-09-360-237-1

Query Match 97.7%; Score 5487; DB 4; Length 1052;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKLVNIMLLLVLLVLCGKHHGLGRLKSEFKAPPCSGSHLLTKVESSVVEVEYVAF 60
Db 1 MKLVNIMLLLVLLVLCGKHHGLGRLKSEFKAPPCSGSHLLTKVESSVVEVEYVAF 60
QY 61 NGYFTAKARNSFTSSALKSEVDNWRITPRNNPSSDYPSPDEVTOIKEROKAGLLTLEDH 120
Db 61 NGYFTAKARNSFTSSALKSEVDNWRITPRNNPSSDYPSPDEVTOIKEROKAGLLTLEDH 120
QY 121 PNIRKRYTPQKRVRSLSKTAESDPYPCNETRMSQKMSRPLKRAASLSLGSGFHNAIGH 180
Db 121 PNIRKRYTPQKRVRSLSKTAESDPYPCNETRMSQKMSRPLKRAASLSLGSGFHNAIGH 180
QY 181 SSRRLKAIPTROVAQTQADVLMOMGTGANVRVAVEDTGLSEKHPHKVKERTMTNTNE 240
Db 181 SSRRLKAIPTROVAQTQADVLMOMGTGANVRVAVEDTGLSEKHPHKVKERTMTNTNE 240
QY 241 RTLDGIGHGTFAVAVIASMKECGFAPDAELHIFRYFTNNQVSYTSMFLDAFNAYAILKK 300
Db 241 RTLDGIGHGTFAVAVIASMKECGFAPDAELHIFRYFTNNQVSYTSMFLDAFNAYAILKK 300
QY 301 IDVNLNLSIGGPDHDFVNDKWMELTANNVIMSAIGNDGLYGLTLNPPADOMVYIGVG 360
Db 301 IDVNLNLSIGGPDHDFVNDKWMELTANNVIMSAIGNDGLYGLTLNPPADOMVYIGVG 360
QY 361 IDEFNTIARFSSRGMETWELPGYGRMKPDIVTYGAGVRSRGVGGCRAISGTSVASPVV 420
Db 361 IDEFNTIARFSSRGMETWELPGYGRMKPDIVTYGAGVRSRGVGGCRAISGTSVASPVV 420
QY 421 AGAVTLIVSVQKRELVNPAKMAQALIASARLPEGVNNEEQGHGKLDLRLAYQILNSYKP 480
Db 421 AGAVTLIVSVQKRELVNPAKMAQALIASARLPEGVNNEEQGHGKLDLRLAYQILNSYKP 480
QY 481 QASISPSYIDLTETCPYMPYCOPITYYGMMTYVNTILNMGYGRIVDKPDMQPIYLPQ 540
Db 481 QASISPSYIDLTETCPYMPYCOPITYYGMMTYVNTILNMGYGRIVDKPDMQPIYLPQ 540

QY 541 NGDNIEAFSSYSVLMWPGSLAISISVTKKAASMEGIAQGHVMTVVASPAETESKNGAE 600
Db 541 NGDNIEAFSSYSVLMWPGSLAISISVTKKAASMEGIAQGHVMTVVASPAETESKNGAE 600
QY 601 QSTVVKLPKVKIIPTPPRSKRVLMDOYHNLRYPPGYPFRDNLKMKNDPLDMNGDHJTN 660
Db 601 HTSTVVKLPKVKIIPTPPRSKRVLMDOYHNLRYPPGYPFRDNLKMKNDPLDMNGDHJTN 660
QY 661 FRDMYQHLRSMGVVEVVLGAPFTCFDASQYOTLLMNDSEEEYFPEELAKLRDNDGLSL 720
Db 661 FRDMYQHLRSMGVVEVVLGAPFTCFDASQYOTLLMNDSEEEYFPEELAKLRDNDGLSL 720
QY 721 VIFSDWNTSVMRKVFYDETRMMPDGTGAGANIPALNELLSVNMNGFSGLYEGEFTL 780
Db 721 VIFSDWNTSVMRKVFYDETRMMPDGTGAGANIPALNELLSVNMNGFSGLYEGEFTL 780
QY 781 ANHDMYASGCSIAKFPEDGVITQTEKDOGLEVKOETAVENVPIGLYQIPAEAGGR 840
Db 781 ANHDMYASGCSIAKFPEDGVITQTEKDOGLEVKOETAVENVPIGLYQIPAEAGGR 840
QY 841 ILYGDSNCLDSDHROKDCFVLDLALQYTSYGYTPPSLSHSGNRORPPSGAGSVTERM 900
Db 841 ILYGDSNCLDSDHROKDCFVLDLALQYTSYGYTPPSLSHSGNRORPPSGAGSVTERM 900
QY 901 EGNHLHRSKYLEAHLGDPKRPPLPACPRLSMAKPOPENETAPSNLTKHOKLSIDLDKY 960
Db 901 EGNHLHRSKYLEAHLGDPKRPPLPACPRLSMAKPOPENETAPSNLTKHOKLSIDLDKY 960
QY 961 VLPNFRSNRPQVRLSPGSGAMDPGIMGRYNOEVGOTIPVAFAGAVYLAFFVQ 1020
Db 961 VLPNFRSNRPQVRLSPGSGAMDPGIMGRYNOEVGOTIPVAFAGAVYLAFFVQ 1020
QY 1021 INKAKSRKRRKPRVKRPOLMOQVHPKPTSV 1052
Db 1021 INKAKSRKRRKPRVKRPOLMOQVHPKPTSV 1052

RESULT 4

US-07-923-260A-6
Sequence 6, Application US/07923260A
Patent No. 5719021
GENERAL INFORMATION:
APPLICANT: Inouye, Masaori
TITLE OF INVENTION: PROTEIN ACTIVATION
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gerard J. Weiser
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/923, 260A
FILING DATE: 31-JUL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 377,5638P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid
TOPOLOGY: linear

Mon Jun 2 11:46:23 2003

us-09-830-837-6.rai

Page 4

```

1 MOLECULE TYPE: protein
2 HYPOTHETICAL: NO
3 ANTI-SENSE: NO
4 ORIGINAL SOURCE:
5 ORGANISM: Bacillus subtilis
6 STRAIN: 168
7
8 FEATURE:
9 NAME/KEY: Region
10 LOCATION: 1..77
11 OTHER INFORMATION:
12 OTHER INFORMATION:
13
14 FEATURE:
15 NAME/KEY: Region
16 LOCATION: 18
17 OTHER INFORMATION:
18 OTHER INFORMATION:
19
20 FEATURE:
21 NAME/KEY: Region
22 LOCATION: 265
23 OTHER INFORMATION:
24 OTHER INFORMATION:
25
26 /note="The region from 1 to 77 is
27 a propeptide."
28
29 /note="Amino acid residue 18 is
30 methionine or threonine."
31
32 /note="Amino acid residue 265 is
33 serine or leucine."
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010

```

Query Match	6.08;	Score 338.5;	DB 1;	Length 352;
Best Local Similarity	26.88;	Pred. No. 3.4e-19;		
Matches 117; Conservative	66;	Mismatches 156;	Indels 97;	Gaps 177;

```

OY      48 SSVVEVEYIYAENGIFTAKRNSPSSALKSSEVDNWRRIIPRNPSSDYPSPDEVIQIK 107
        ' III I : III I : III I :
Db      4 SST--EKKYIVG-----KQTXSAMSSAKKDYI----- 30

```

```

QY      108 EKOKAGLLTLE-DHPNIKRPTQKRVFRSLKYAESDPYPCPNETRMSQKWOSSRPLRRAS 166
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      31 -SEKGVQKQPKYVNAATAATLDEKAVKEIK--KPSVA----- 66

```

```

QY      167 LSLGSGFWHAICRHHSSRLLRAIPROVAQLQADVLDMQGYGTGANVRVAVEFDGLSEKHP 226
          :      |      :      :      :      :      :      :      :      :      :
Db      67 -----YVEEDHIAHEYAQSVPYGISQ- IKAPALHSQGYTSGSNVKAIVADISGIDSSHP 117

```

0y 227 HFKNVRKRTINTNTNERT--LDDGLGCHGTFAVGVIASMR---CGGFAPDAELHIFRYPTNN 281
 118 DL-NVRGASFPSEINPYDGGSHGTHVAGTIALNNSIGVLGVSPASLYAVKVLDTST 176

262 QVSTISAELEDAENYALFLKKIDVLNLSIGGDFMD--HPYVDKWEFTANNVNVSAIGND 339
 || :: :||:::|:|:| |||
 Db 177 GSGQYSHLINGIEMALSNNDVYNNSLGGPTGSLTKTYDKA--YSSGIYVAAAAGNE 233

234 GSSGSGTGVGPAPKYPSTIAIVGAVNSSNQNRKAFSSAGS---EL-----DYMAPGVS 281 Db

Db

282 IOSTLPGGTGAYNGTSMATPHVAGAAALILS---KHPMTNAQVRDLRESTATYLGS-N 336

Db 337 SFYGGGLINVOAAQ 352

RESULT 5
US-07-923-260A-1
: Sequence 1, Application US/07923260A

```

: GENERAL INFORMATION:
:
: APPLICANT: Inouye, Masayori
: TITLE OF INVENTION: PROTEIN ACTIVATION
: NUMBER OF DRAWINGS:

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

STATE: PA

COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/923,260A
FILING DATE: 31-JUL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weisler, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 377,5638P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8334
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid

MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO

ORGANISM: *Bacillus subtilis*
STRAIN: 168
FEATURE:

```
LOCATION: 1.77  
OTHER INFORMATION: /note="The region from 1 t  
OTHER INFORMATION: a propeptide."
```

Query Match	6.08;	Score 337.5;	DB 1;
Best Local Similarity	26.88;	Pred. No. 4.1e-19;	
Matches 117.	Conservative	66.	Mismatches 156.

QY 48 SSTVVEYEYIVAFNGYFTAKARNSPSSALKSGSEVDNMRIL
||| : ||| | : ||| :
Db 4 SST--FKKYTVGF-----KOTMSAANSAKFKDVT-----

QY 108 EKAKAGLTLE-DHPNIRKVTPOKRVFSLKLKVAESDPVVF
:
| : : : | : : : | : : : | : : : | : : : | : : :
Db 31 -SEGGKVKOKFKVNAAATLDEKAVELK---KDPSTVA

```

QY      167 LSLGSGFWHATGHRSSRLRLRAPROVAOTLOADVYLMKONG
        : | : :: : | : | |
Db      67 -----YVEEDINAEYAQSPYGISQ-IKAPALHSQG

```

```

QY      227 HEKVKERNWTNERT--LDDGLGHTFAGVATASMR-----
      11:  ::  -  11 111 111 11:
Db      118 DL-NVRGGASFVSETPYQDGGSHGHTAGTAAALNNSI

```

```
QY      282 QVSYTSWFLDAFNVAILEKKIDVNLNLSIGSGPDFMD--HDFV  
          ||::|||::|:|:|:|:|:|:|_  
Db     177 GSGCYSWINGIEFAISNMNDVYNMSLGGPTGSAALKTVV
```

QY 340 GP--LYGTLLNPADQMDVIGVGIDFEDINARFSSRGMTT
 Db 224 GSSGSTSYVGPKYPSTIATGAVNSSNRASSTAGS---

OY 398 VRSGGVKGGCRALSTSVASPVVAGAVTLLVSFQVKRELIV
:::|::|||:|::|
Db 282 IQGTLPGGTGYGAINGTSMATPHVGAALILS---KHPTW

QY 458 MFEGHGKLDLRLAYQ 473
| | | :: | |
Db 337 SFYYGKGLINQAAQ 352

17.

----- 30

RRAS 166
----- 66

SEKHP 226
||
SSHP 117

ETNN 281
:
LDST 176

```

.IGND 339
  ||:
  AGNE 233

```

Y G A G 397
|
P G V S 281

PGVN 457
1
G--N 336

RESULT 6

US-07-772-087-5
: Sequence 5, Application US/07772087
: Patent No. 5275945
: GENERAL INFORMATION:
: APPLICANT: HSIAO, Hung-Yu
: APPLICANT: FODGE, Douglas W.
: APPLICANT: LALONDE, James J.
: TITLE OF INVENTION: ALKALINE PROTEASES STABLE IN HEAVY-DUTY
: TITLE OF INVENTION: DETERGENT LIQUIDS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: City: Alexandria
: State: Virginia
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/772,087
: FILING DATE: 19911008
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 16754/115 CHCO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 381 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Bacillus
: STRAIN: B. subtilis
: US-07-772-087-5

Query Match 6.0%; Score 337.5; DB 1; Length 381;
Best Local Similarity 26.8%; Pred. No. 4.6e-19;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSTVEVEYIVANNGFTAKARNSFTSSALKSSSEYDWMKRIIPRNPSSDIPSPDEVIQIK 107
DB 33 SST-EKRTIVGF-----KOTMSAMSSAKKKDVI----- 59
QY 108 EKOKAGILTLE-DHPNIKRYTPQKRYFRSLKYAESDPTVPCNETRMSQKMSRPLRRAS 166
DB 60 -SEKGGVQKQFYVNAANAATLDEKAYKELK---KDSVA----- 95
QY 167 LSLGSGFWATGRRSSRLRLRAIPROYAQTLDVLMQMGYTGANVAVFDTGLSEKHP 226
DB 96 -----YEEEDHIAHEYAOSVPGYSQ- IKAPALHSQGYTGSNVVAVVYDSDIDSHP 146
QY 227 HFKNVKERTWMTNERT--LDDGLGHGTFVAGVIASME---COGFADALHIFRYFTNN 281
DB 147 DL-NVRGASFEVSETPYQDSSGHGTHVAGTIALNNSIGVLGVSFASLVAVKYLDST 205
QY 282 QVSTSWFLDAFNVAIILKIDVNLSTIGGDFMD--HPFYDKWELTANNVIMVSAIGND 339
DB 206 GSGQYIMILINGIMIALSNMNDVIMSLGPTGSTALKTVYDKA---VSSGIIVAAAGNE 262
QY 340 GP--LYGTLLNPADQMDVIGVGIGIDEDNIARFSSRQMTWELPGYGRKKPDIYTYGAG 397
DB 263 GSGSGSTVGVYPAKYPTSTIAVGAVNNSNQASFSAGS---EL-----DYMAFGVS 310

QY 398 VRSGVYGGCRALSGTSVAPVAGATYLLVSTYQKREIYNPASMKQALLSARRLPGVN 457
DB 311 IQSTLPQGYGAYNGTSMATPHVAGAAALILS---KHPTWTNAOVRDLSEATYILG--N 365
QY 458 MEEQHGKLDLLRAYQ 473
DB 366 SFYFGKGLINVOAAQ 381

RESULT 7

US-08-322-965-2
: Sequence 2, Application US/08322965
: Patent No. 5733473
: GENERAL INFORMATION:
: APPLICANT: Johnston, James P.
: APPLICANT: Lenoir, Pierre M.A.
: APPLICANT: Thoen, Christian A.J.K
: TITLE OF INVENTION: Liquid detergent Composition containing
: TITLE OF INVENTION: lipase and Protease
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: The Procter & Gamble Company
: STREET: 5299 Spring Grove Avenue
: City: Cincinnati
: State: Ohio
: COUNTRY: USA
: ZIP: 45217
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/322,965
: FILING DATE: 13-OCT-1994
: CLASSIFICATION: 252
: ATTORNEY/AGENT INFORMATION:
: NAME: Allen, George W.
: REGISTRATION NUMBER: 26,143
: REFERENCE/DOCKET NUMBER: CM-352MC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 513/627-5946
: TELEFAX: 513/627-8118
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 381 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-322-965-2

Query Match 6.0%; Score 337.5; DB 1; Length 381;
Best Local Similarity 26.8%; Pred. No. 4.6e-19;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSTVEVEYIVANNGFTAKARNSFTSSALKSSSEYDWMKRIIPRNPSSDIPSPDEVIQIK 107
DB 33 SST-EKRTIVGF-----KOTMSAMSSAKKKDVI----- 59
QY 108 EKOKAGILTLE-DHPNIKRYTPQKRYFRSLKYAESDPTVPCNETRMSQKMSRPLRRAS 166
DB 60 -SEKGGVQKQFYVNAANAATLDEKAYKELK---KDSVA----- 95
QY 167 LSLGSGFWATGRRSSRLRLRAIPROYAQTLDVLMQMGYTGANVAVFDTGLSEKHP 226
DB 96 -----YEEEDHIAHEYAOSVPGYSQ- IKAPALHSQGYTGSNVVAVVYDSDIDSHP 146
QY 227 HFKNVKERTWMTNERT--LDDGLGHGTFVAGVIASME---COGFADALHIFRYFTNN 281
DB 147 DL-NVRGASFEVSETPYQDSSGHGTHVAGTIALNNSIGVLGVSFASLVAVKYLDST 205
QY 282 QVSTSWFLDAFNVAIILKIDVNLSTIGGDFMD--HPFYDKWELTANNVIMVSAIGND 339

DB 206 GSGQSVMIINGIENAMISNNNDVIMNLSGPGTSTALKTVDKA---VSSGIYVAAAANE 262
OY 340 GP--LXGLINPADOMOVIGVGGIDEDNIAFSSRGTTWELPGYGRMKPDITYAG 397
DB 263 GSSGSTIYGPAPKPTSTIAGVAVNSNORASFSSAGS---EL-----DYAPGV 310
OY 398 VRSQYKGGCRLAGSTVSPVAVGAVTLVSTVOKRELVPASAKOALIASARLPV 457
DB 311 IOSTLPGYKINGTSMATPHVAGAAALITS---KHPTWNAOVRDLSTATYLG--N 365
OY 458 MFEQHGKIDLRAYQ 473
DB 366 SFYGGKGLINVOAAQ 381

RESULT 8

US-07-864-298-1
Sequence 1, Application US/07864298
Patent No. 5316935
GENERAL INFORMATION:
APPLICANT: Arnold, Frances H.
TITLE OF INVENTION: No. 5316935el Subtilisin Variants
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harris F. Brotman, PQMS, SMITH, LANDE & ROSE
STREET: 2121 Avenue of the Stars Suite 1400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOOS
SOFTWARE: Patent release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/864, 298
FILING DATE: 1992/04/06
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Brotman, Harris F.
REGISTRATION NUMBER: 35,461
REFERENCE/DOCKET NUMBER: 91-124
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus subtilis
STRAIN: 1168
US-07-864-298-1

Query Match

Best Local Similarity 5.9%; Score 330.5; DB 1; Length 275;
Best Local Similarity 31.8%; Pred. No. 1e-18;
Matches 94; Conservative 54; Mismatches 117; Indels 31; Gaps 11;
DB 187 RAIPROVQTLQADVIMQMGYTGAVRVAVDTGLSEKHFKNVETRTNMTNE--LD 244
OY 2 OSVPFGISQ--IKAPALHSQGYTSKNKAVYDSDGSDSHDL--NVGSGASFPSETNPYQ 59
DB 245 DLGAGTGVACVYASME--COGEPADAEHLHFRVFTNNQVSTYSMPFLDAFNAILKKI 301
OY 60 DSSHGTHVACTIALNNSIGVYVAPASASLYAVKVLDSGQSVMIINGIENAMISNM 119

OY 302 DYLANISIGGDFEMD--HPPYDKWEILANNVIMVSAIGNDGP--LYGLINPADOMDYG 357
DB 120 DYKMSLGGTGTSTALKTYVDKA---VSSGIYVAAAAGNESSGSIYSTIYGPAPKPTSTIA 176
OY 358 VGGIDFEDNIAFSSRGTTWELPGYGRMKPDITYAGVRSQYKGGCRLAGSTV 417
DB 177 VGANVSNORASFSSAGS---EL-----DYAPGVSTOSTLPGYKINGTSMAT 224
OY 418 PVVAGAVTLVSTVOKRELVPASAKOALIASARLPVNNFEQHGKIDLRAYQ 473
DB 225 PHVAGAAALITS---KHPTWNAOVRDLSTATYLG--NSFYGGKGLINVOAAQ 275

RESULT 9

US-08-845-295A-1
Sequence 1, Application US/08845295A
Patent No. 5817490
GENERAL INFORMATION:
APPLICANT: Hubbs, John C.
TITLE OF INVENTION: Enzymatic Process for the Manufacture of
TITLE OF INVENTION: Ascorbic Acid, 2-Keto-L-Gulonic Acid, and Esters of
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eastman Chemical Company
STREET: P. O. Box 511
CITY: Kingsport
STATE: Tennessee
COUNTRY: USA
ZIP: 37662-5075
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch disk
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845, 295A
FILING DATE: 25-April-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017, 879
FILING DATE: 17-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Cheryl J. Tubach
REGISTRATION NUMBER: 38,346
REFERENCE/DOCKET NUMBER: 70432
TELECOMMUNICATION INFORMATION:
TELEPHONE: 423-229-6189
TELEFAX: 423-229-1239
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-845-295A-1

Query Match

Best Local Similarity 5.9%; Score 329.5; DB 2; Length 379;
Best Local Similarity 25.3%; Pred. No. 2e-18;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;
DB 1 MKLVNMLLVLLVLCGKKHIGDLREKKSFERAPCGGSHLTILKVEFSSTV-----V 52
OY 2 MRKSPFLGTLAFM-----LVFTMAFSDSASAPAKNV 36
DB 53 EYEYIAF--NGYFTAKANSPFISALKSSEVD--NMRLIPRNNSSDYPSDFEVIOIREKQ 110
OY 37 EKDYIYGFSGVGYKATVSKKLIKES--GGKVDKOPRIINAKAKALD-----80
DB 111 KAGLTLEHPINIKVTPORVFRSLKYAESDPYPCNERTFMSQKQSSRPLRASIIG 170
OY 81 KEALKEVKNDPPV-----AYVEED-----99

```

QY 171 SCFMATGRSHSRRLRLAPROVAOTGLADLYLWOMQYUANTANVAYAEJDSKHPFEKN 230
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 100 -----HAHALAAGTYPGCI -PLIKADKYQAQGFKANKANVAILDITGICASHEDLVN 149
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 VKERTNMTNERTLDGLGHGTFFVAGVIASMRQO--GAPDAELHIFVFTNNQVSYS 287
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 VGASFFVAGVEAYNTGNGCHGTFFVAGVVAALDNTTGLVLAAPSVSLAAYKLNSSGSGTYS 209
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 WFIDAFNAILKIDIVLNLISIGPPEMD--HPFVDYKWEILTANVIMSAIGNDGPL- Y 343
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 210 GIVSIEBATTNGMIVNINSLGSPGSGTLMKQAVNDAY---ARGVYVVAALAGSGSGSNT 266
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 GTLNNPADQMDVYIGVGIDFEDNIAFFSSRGMTWELPGYGIMKRPDIYTYGAGVNGSGV 403
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 NTIGYPAYDYSVIAAGVADSNSNRASFSSVGA---EL-----EVNAPGAGVYSTYP 314
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 404 KGGCRALSGTSVASFVAVGAVYLLVSTVQKRELVNPSAKOALIASARRLPGVNMFEQGH 463
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 TSYATLNGTSMASPHVAGAAALITLS--KHBNLSASQVRRNRLRSATYTLG--SSFYEGK 369
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 464 GKLDLLRAYQ 473
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 GLINVEAAAO 379
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 1 KLUVNIWLLLVYLLQCKHKLHGBORLEKKSFEKAPCPSCMLLLKVESSVY-----Y 52
   | : : : : : |
Db 2 MRKSFWDGMLTAFM-----LVFTMAFSDSASAAOPKAV 36
   | : : : : : |
QY 53 EYEXIVAF-NGYFTPAKARNSPISALKSSEVD--NMRIIPRNPPSSDYPSDFEYIOIKKO 110
   | : : : : : |
Db 37 EKDIYVGFKSGVKTAASYKDIKES--GKVKDQOFRIINAKAKALD-----80
   | : : : : : |
QY 111 KAGLLTLEDDHPNIKRYVPQKRVPSLKYAESDPYVPCNEIRMQKMSRPLKRAISLIG 170
   | : : : : : |
Db 81 KEALKEVKNPDV-----AYEBD-----99
   | : : : : : |
QY 171 SGFWHAYGRHSRRLTALAIROYAQLQADAVLMQGTGANVRYAVEDTGLSEKHPIFKN 230
   | : : : : : |
Db 100 -----HVAHMLAQTPVYGI-PLIKADKVLQAGSKGANVKAVALDGTQIASHPOLAV 149
   | : : : : : |
QY 231 VKERTMNTNETLDDGIGCHGTFYAGVAGIYASBREC-----GFPADDELHIFRFTNNQVSYTS 287
   | : : : : : |
Db 150 VVGASFYVAGEVYNTDNGHGHVAGTVAALDNTTGVIGVAPSVSLVAVKVLSSGSGSTYS 209
   | : : : : : |
QY 288 WFLDAFYAYALIKKIDVNLNLSIGGPFMD--HPFYDKWELTANNVNVSAIGNDGPL-X 343
   | : : : : : |
Db 210 GIVSGIEMATNGMDVIMNSLIGSPSGSTAMKQAVDNY--ARGVYVAAAAGNSGSGSNT 266
   | : : : : : |
QY 344 GTLNNPDDQMDYGVGGIDFEDNFIARRSSKGMTWELPGVGRKKPDIVYVYGVGRSSGV 403
   | : : : : : |
Db 267 NFIQYPAKYDSVIAVGAVDNSNRPASSVSGA--EL-----EVMAPGAGVSTYP 314
   | : : : : : |
QY 404 KGGCRALSGTSVASPVVAVGAVTLLVSTVQJRELNVPAISMKQALIASARLDPGVNMPEOGH 463
   | : : : : : |
Db 315 TSTVATLNGTSMASPHVAGAAALILS--KHPNLASAQVNRRLSSTATYLG--SSFYYK 369
   | : : : : : |
QY 464 GKLDLLRAYQ 473
   | : : : : : |
Db 370 GLINVEAAQ 379

```

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-09-146-661-1

Query Match 5.98; Score 329.5; DB 4; Length 379;
Best Local Similarity 25.38; Pred. No. 26-18;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;

1 MKLVNIMLLLVLLVLLCGKKHLDRLKSKFEKAPCGCSHLTLKVEFSSTV-----V 52
2 MKKSPFWGLMLTAFM-----LVFTMAFSDSASAAQPAKNV 36
3 EYEYIAF-NGYFTAKARNSTFSSALKSEVD-NMRTIPRNPSDYPDEYIOIKKQ 110
4 EKDYIVGFKSGVTKASVKDLIKES--GGKVDKQRIINAKAKLD----- 80
5 37 KAGLTLEDPNIRKRVTPQKRVRSLLKVAESDPYPCNETRMSQKQSSRPLRRASLSLG 170
6 81 KEALKEKNDPV-----AYEED----- 99
7 171 SGFWHATGRSSRLRLAIPROVAQTLOADVIMQGYTGANRVAVFDGLSEKHPRKN 230
8 100 -----HVAHALAQVTPYGI-PLIKADKVOAGFGANVAVLDYGIASHPDLNV 149
9 231 VKERTNMTNERTLDGLGHTGFVAGVIASMRQCO--GFAPDELHIFRYFTNNQVSYTS 287
10 150 VGASGVAGVAVNTDNGHGTHTVAGTVALDNTGTGLVAPSVSLAVVYVLLSSSGSYTS 209
11 288 WFLDAFNVAITLKIDVNLISIGDPFMD--HPFVDKWELETANNVIMVSAIGNDPPL--Y 343
12 210 GIVSGIEMATTNGMDVIMSLGPGSGSTAMKQAVDNAY--ARGVYVVAAGNSSSGNT 266
13 344 GTLNNPADQMDVIGVIGIDFEDNIARFSSRGMTWELPGYGRMKPDITYGAGVRSQV 403
14 267 NTIGYPAKYDSYIANGVAVDANSNRASFSSVGA--EL-----EVAPAGVYSTYP 314
15 404 KGGCRALSGTSVAPVYAGATLLVSTVQKRELVPASMKOALIASRRLLPVNMFEDGH 463
16 315 TSTYATLNTGTSMAHPVAGAAALILS---KHPNLASQVRNRLSSTATYLG--SSFTYIK 369
17 464 GKLDLIRAYQ 473
18 370 GLINVEAAQ 379

RESULT 12
US-09-150-515-1
Sequence 1, Application US/09150515
Patent No. 6271006

GENERAL INFORMATION:
APPLICANT: Hubbs, John C.
TITLE OF INVENTION: Enzymatic Process for the Manufacture of
TITLE OF INVENTION: Ascorbic Acid, 2-Keto-L-Gulonic Acid, and Esters of
TITLE OF INVENTION: 2-Keto-L-Gulonic Acid
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eastman Chemical Company
STREET: P.O. Box 511
CITY: Kingsport
STATE: Tennessee
COUNTRY: USA
ZIP: 37662-5075
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/150, 515

FILING DATE: 09-SEP-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017,879; 08/845,295
FILING DATE: 17-May-96; 25-April-97
ATTORNEY/AGENT INFORMATION:
NAME: Cheryl J. Tubach
REGISTRATION NUMBER: 38,346
REFERENCE/DOCKET NUMBER: 70432
TELEPHONE: 423-229-6189
TELEFAX: 423-229-1239
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-09-150-515-1

Query Match 5.98; Score 329.5; DB 4; Length 379;
Best Local Similarity 25.38; Pred. No. 26-18;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;

1 MKLVNIMLLLVLLVLLCGKKHLDRLKSKFEKAPCGCSHLTLKVEFSSTV-----V 52
2 MKKSPFWGLMLTAFM-----LVFTMAFSDSASAAQPAKNV 36
3 EYEYIAF-NGYFTAKARNSTFSSALKSEVD-NMRTIPRNPSDYPDEYIOIKKQ 110
4 EKDYIVGFKSGVTKASVKDLIKES--GGKVDKQRIINAKAKLD----- 80
5 37 KAGLTLEDPNIRKRVTPQKRVRSLLKVAESDPYPCNETRMSQKQSSRPLRRASLSLG 170
6 81 KEALKEKNDPV-----AYEED----- 99
7 171 SGFWHATGRSSRLRLAIPROVAQTLOADVIMQGYTGANRVAVFDGLSEKHPRKN 230
8 100 -----HVAHALAQVTPYGI-PLIKADKVOAGFGANVAVLDYGIASHPDLNV 149
9 231 VKERTNMTNERTLDGLGHTGFVAGVIASMRQCO--GFAPDELHIFRYFTNNQVSYTS 287
10 150 VGASGVAGVAVNTDNGHGTHTVAGTVALDNTGTGLVAPSVSLAVVYVLLSSSGSYTS 209
11 288 WFLDAFNVAITLKIDVNLISIGDPFMD--HPFVDKWELETANNVIMVSAIGNDPPL--Y 343
12 210 GIVSGIEMATTNGMDVIMSLGPGSGSTAMKQAVDNAY--ARGVYVVAAGNSSSGNT 266
13 344 GTLNNPADQMDVIGVIGIDFEDNIARFSSRGMTWELPGYGRMKPDITYGAGVRSQV 403
14 267 NTIGYPAKYDSYIANGVAVDANSNRASFSSVGA--EL-----EVAPAGVYSTYP 314
15 404 KGGCRALSGTSVAPVYAGATLLVSTVQKRELVPASMKOALIASRRLLPVNMFEDGH 463
16 315 TSTYATLNTGTSMAHPVAGAAALILS---KHPNLASQVRNRLSSTATYLG--SSFTYIK 369
17 464 GKLDLIRAYQ 473
18 370 GLINVEAAQ 379

RESULT 13
US-07-923-260A-4
Sequence 4, Application US/07923260A
Patent No. 5719021

GENERAL INFORMATION:
APPLICANT: Inouye, Masayori
TITLE OF INVENTION: PROTEIN ACTIVATION
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gerard J. Weiser
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia

OY 429 STYOKRELVPASMKQALLIASARRLPVNMFEQHGKIDLLRAYQ 473
DB 339 S---KHPNLSASOVNRNLSSTATYLG--SSFYGGKGLINVEAQAQ 378

RESULT 15

US-08-322-677A-8
; Sequence 8, Application US/08322677A
; Patent No. 5677272
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Chanchal K.
; APPLICANT: Burns, Michael E.
; APPLICANT: Digulio, David N.
; APPLICANT: Getty, Edward E.
; APPLICANT: Hartshorn, Richard T.
; APPLICANT: Willey, Alan D., III
; APPLICANT: Brode, Philip F., III
; APPLICANT: Barnett, Bobby L.
; APPLICANT: Rubingh, Donn N.
; TITLE OF INVENTION: Bleaching Compositions Comprising Protease Enzymes
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East River Road
; CITY: Cincinnati
; STATE: OH
; COUNTRY: USA
; ZIP: 45253-8707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/322,677A
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Zerby, Kim William
; REGISTRATION NUMBER: 32,323
; REFERENCE/DOCKET NUMBER: 5041R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-2885
; TELEFAX: (513) 627-0318
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-322-677A-8

Query Match 5.8%; Score 323.5; DB 1; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

OY 187 RAIPROVAQTLDADVLQMGGYTGANRVAVFDITGLSEKHPHFKNVERTNMTNERT--LD 244
DB 2 OSVPYGISQ-IRKAPALHSQGYTGSNKVAVIDSGIDSSHDL-NVKGGSFVPSSTNPYQ 59
OY 245 DGLGHTGFVAVIASMR--COGFAPDAELHIFVFTNNQVSTMPFDANVYALKKI 301
DB 60 DSSHGTHVAGTIALNNSIGVLGVSASLYAVVLDSTGSGQYSMIINGIEMALSNM 119
OY 302 DVNLSTISGPDPM--HPFVDKVMELTANNVIMVSAIGNDP--LVGTLANPADQMDVIG 357
DB 120 DVINNSLGGPTGSLTKTVYDKA---VSSGIYVAAAAGNBSGSGYSTYGYAKYPTSTIA 176
OY 358 VGGIDFEDNINARSSRGKTTWELPGGYGKMKPDYTYTGAVRSGSYKGGCRALSGTSTYAS 417
DB 177 VGAVNNSNORASFSASG--EL-----DVNAPEVSTIGTLPGGTYGAYNIGTSMAT 224

OY 418 PVVAGAVTLVSTYOKRELVPASMKQALLIASARRLPVNMFEQHGKIDLLRAYQ 473
DB 225 PVVAGAAALITL---KHPNLSASOVNRNLSSTATYLG--NSFYGGKGLINVOAQAQ 275

RESULT 16

US-08-322-676-8
; Sequence 8, Application US/08322676
; Patent No. 5679630
; GENERAL INFORMATION:
; APPLICANT: Baeck, Andre (NMN)
; APPLICANT: Ghosh, Chanchal K.
; APPLICANT: Graycar, Thomas P.
; APPLICANT: Bott, Richard R.
; APPLICANT: Wilson, Lori J.
; APPLICANT: Brode, Philip F., III
; APPLICANT: Barnett, Bobby L.
; APPLICANT: Rubingh, Donn N.
; TITLE OF INVENTION: Protease-Containing Cleaning Compositions
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East River Road
; CITY: Cincinnati
; STATE: OH
; COUNTRY: USA
; ZIP: 45253-8707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/322,676
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Zerby, Kim William
; REGISTRATION NUMBER: 32,323
; REFERENCE/DOCKET NUMBER: 5040R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-2885
; TELEFAX: (513) 627-0318
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-322-676-8

Query Match 5.8%; Score 323.5; DB 1; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

OY 187 RAIPROVAQTLDADVLQMGGYTGANRVAVFDITGLSEKHPHFKNVERTNMTNERT--LD 244
DB 2 OSVPYGISQ-IRKAPALHSQGYTGSNKVAVIDSGIDSSHDL-NVKGGSFVPSSTNPYQ 59
OY 245 DGLGHTGFVAVIASMR--COGFAPDAELHIFVFTNNQVSTMPFDANVYALKKI 301
DB 60 DSSHGTHVAGTIALNNSIGVLGVSASLYAVVLDSTGSGQYSMIINGIEMALSNM 119
OY 302 DVNLSTISGPDPM--HPFVDKVMELTANNVIMVSAIGNDP--LVGTLANPADQMDVIG 357
DB 120 DVINNSLGGPTGSLTKTVYDKA---VSSGIYVAAAAGNBSGSGYSTYGYAKYPTSTIA 176
OY 358 VGGIDFEDNINARSSRGKTTWELPGGYGKMKPDYTYTGAVRSGSYKGGCRALSGTSTYAS 417
DB 177 VGAVNNSNORASFSASG--EL-----DVNAPEVSTIGTLPGGTYGAYNIGTSMAT 224

Mon Jun 2 11:46:23 2003

us-09-830-837-6.ra1

Page 12

OY 358 VGGIDFEDNINARFSSRGMTWELPGCYGRMKPDIYTYGAGVSGVKGCGALSGTSVAS 417
Db 177 VGVNSSNRASFSAGS---EL-----DWMAGVSIQSTLPGGTGAYNGTSMAT 224
OY 418 PVVAGAVTLVSTVQKRELVPASMKOALIASARLPGVMMFQGHGKLDLIRAYQ 473
Db 225 PVVAGAAALILS---KHPTWNAQVRDLRESTATYILG--NSFYFGKGLINVOAAQ 275

RESULT 19
US-09-255-502-3
Sequence 3, Application US/09255502
Patent No. 6218165
GENERAL INFORMATION:
APPLICANT: Estell, David
APPLICANT: Harding, Fiona
TITLE OF INVENTION: Mutant Proteins Having Lower Allergenic Responses in
TITLE OF INVENTION: Humans and Methods for Constructing, Identifying and
TITLE OF INVENTION: Producing Such Proteins
FILE REFERENCE: GC 527-D2
CURRENT FILING DATE: 1998-02-23
PRIOR FILING DATE: 1998-04-15
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 3
LENGTH: 275
TYPE: PRT
ORGANISM: Bacillus subtilis
US-09-255-502-3

Query Match 5.8%; Score 323.5; DB 4; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

OY 187 RAIPROVACTLOADVLNOMQYTGANVRVAVFTGLSEKHPKKNKERTNMTNERT--LD 244
Db 2 QSVYVGISQ--IKAPLHSGCYTSNKKVAVIDSGIDSHPD--NVRGASFPVSETPYQ 59
OY 245 DGLGHTFVAGVIASMR---COGFAPDAELHIFRFTNNQVSTSWFLDAFNAILKRI 301
Db 60 DGSCHGHVAGTIALNNSIGVSPSASLYAVLIDSTGSGYSWIINGIEALISNM 119
OY 302 DVNLSTIGGPDMD--HPEVDKWEELTANNVIMVSAIGNDP--LYGTLPNPDQMDVIG 357
Db 120 DVNMSLGGPTGTALKTVDKA---VSSGIYVAAAAGNBSGSGSTGYGPAKYPSTIA 176
OY 358 VGGIDFEDNINARFSSRGMTWELPGCYGRMKPDIYTYGAGVSGVKGCGALSGTSVAS 417
Db 177 VGVNSSNRASFSAGS---EL-----DWMAGVSIQSTLPGGTGAYNGTSMAT 224
OY 418 PVVAGAVTLVSTVQKRELVPASMKOALIASARLPGVMMFQGHGKLDLIRAYQ 473
Db 225 PVVAGAAALILS---KHPTWNAQVRDLRESTATYILG--NSFYFGKGLINVOAAQ 275

RESULT 20
US-09-178-155-4
Sequence 4, Application US/09178155
Patent No. 6312936
GENERAL INFORMATION:
APPLICANT: Poulos, Aycockaran J.
APPLICANT: Schellenderger, Volker
APPLICANT: Kellis, Jr., James T.
APPLICANT: Paech, Christian
APPLICANT: Nader, Joanne
APPLICANT: Naki, Donald P.
TITLE OF INVENTION: Multiply-Substituted Protease Variants
FILE REFERENCE: GC502-2
CURRENT FILING DATE: 1998-10-23

EARLIER APPLICATION NUMBER: 08/956,323
EARLIER FILING DATE: 1997-10-23
EARLIER APPLICATION NUMBER: 08/956,564
EARLIER FILING DATE: 1997-10-23
EARLIER APPLICATION NUMBER: 08/956,324
EARLIER FILING DATE: 1997-10-23
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 275
TYPE: PRT
ORGANISM: B. subtilis
US-09-178-155-4

Query Match 5.8%; Score 323.5; DB 4; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

OY 187 RAIPROVACTLOADVLNOMQYTGANVRVAVFTGLSEKHPKKNKERTNMTNERT--LD 244
Db 2 QSVYVGISQ--IKAPLHSGCYTSNKKVAVIDSGIDSHPD--NVRGASFPVSETPYQ 59
OY 245 DGLGHTFVAGVIASMR---COGFAPDAELHIFRFTNNQVSTSWFLDAFNAILKRI 301
Db 60 DGSCHGHVAGTIALNNSIGVSPSASLYAVLIDSTGSGYSWIINGIEALISNM 119
OY 302 DVNLSTIGGPDMD--HPEVDKWEELTANNVIMVSAIGNDP--LYGTLPNPDQMDVIG 357
Db 120 DVNMSLGGPTGTALKTVDKA---VSSGIYVAAAAGNBSGSGSTGYGPAKYPSTIA 176
OY 358 VGGIDFEDNINARFSSRGMTWELPGCYGRMKPDIYTYGAGVSGVKGCGALSGTSVAS 417
Db 177 VGVNSSNRASFSAGS---EL-----DWMAGVSIQSTLPGGTGAYNGTSMAT 224
OY 418 PVVAGAVTLVSTVQKRELVPASMKOALIASARLPGVMMFQGHGKLDLIRAYQ 473
Db 225 PVVAGAAALILS---KHPTWNAQVRDLRESTATYILG--NSFYFGKGLINVOAAQ 275

RESULT 21
US-09-445-270-3
Sequence 3, Application US/09445270
Patent No. 6369011
GENERAL INFORMATION:
APPLICANT: Rai, Saroj
APPLICANT: Correa, Paul
APPLICANT: Zhu, Yong
APPLICANT: Graycar, Thomas
APPLICANT: Bott, Richard
TITLE OF INVENTION: Protease Enzymes for Tough Cleaning and
TITLE OF INVENTION: Compositions Incorporating Same
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Procter & Gamble Company
STREET: 11810 East Miami River Road
CITY: Ross
STATE: OH
COUNTRY: USA
ZIP: 45061
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA
APPLICATION NUMBER: US/09/445,270
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Bolam, Brian M.
REGISTRATION NUMBER: 37,513
REFERENCE/DOCKET NUMBER: 1234
TELECOMMUNICATION INFORMATION:

TELEPHONE: 513/627-2457
TELEFAX: 513/627-0318
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-445-270-3

Query Match 5.8%; Score 323.5; DB 4; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

QY 187 RAIPROYAOTLQADVLQMWQYTGANRVAVFDGLSEKHFKNVKERTWMTERT--LD 244
Db 2 QSVYPGISQ-IRAPALHSOGYTGSNVAVIDSGIDSSHDL-NVRGASFEVSETPYQ 59
QY 245 DGLGHGTFFVAGVYASMR--COGFAPDAELHIFRVFTNNQVSYTSMFLDAFYALTKI 301
Db 60 DSSSHGTHVAGTTAALNNSIGVLGVSPASLIVAKYLDSTGSGOYSLINGIEMALSNM 119
QY 302 DVNLISIGDPDMD--HPFVDKYWELTANNVIVSAIGNDP--LYGTLNPPADQMDVIG 357
Db 120 DVINMSLGGPTGSTALLKTYVDKA--VSSGIYVAAAAGNCGSSGSTVGYPAKYPSTIA 176
QY 358 VGGIDEDNINARSSRGMTTWELPGYGRMKPDIVYVYAGVRSYKGGCRALSGTSVAS 417
Db 177 VGVAVNSNORASFSAGS--EL-----DVMAPGVSIQSTLPGITGAYGANGTSMAT 224
QY 418 PVVAGAVTLLVSTVOKRELVPASMKOALIASARLPGVMPFGQHGKLDLLRAYQ 473
Db 225 PHVAGAAALILS--KHPTWTNAQVRDLRESTATYIG--NSFYGGKGLINVOAAQ 275

RESULT 22

5472855-4

Patent No. 5472855

APPLICANT: CARTER, PAUL J.; WELLS, JAMES A.

TITLE OF INVENTION: SUBSTRATE ASSISTED CATALYSIS

NUMBER OF SEQUENCES: 31

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/287,964

FILING DATE: 22-SEP-1994

Prior Application DATA:

APPLICATION NUMBER: 90,902

FILING DATE: 12-JUL-1993

APPLICATION NUMBER: 823,039

FILING DATE: 14-JAN-1992

APPLICATION NUMBER: 35,652

FILING DATE: 06-APR-1987

APPLICATION NUMBER: 334,081

FILING DATE: 04-APR-1989

APPLICATION NUMBER: 127,134

FILING DATE: 01-DEC-1987

APPLICATION NUMBER: 846,627

FILING DATE: 01-APR-1986

APPLICATION NUMBER: 858,594

FILING DATE: 30-APR-1986

APPLICATION NUMBER: 614,612

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,615

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,617

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,491

FILING DATE: 29-MAY-1984

SEQ ID NO: 4:

LENGTH: 275

5472855-4

Query Match

5.8%; Score 323.5; DB 6; Length 275;

Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

QY 187 RAIPROYAOTLQADVLQMWQYTGANRVAVFDGLSEKHFKNVKERTWMTERT--LD 244
Db 2 QSVYPGISQ-IRAPALHSOGYTGSNVAVIDSGIDSSHDL-NVRGASFEVSETPYQ 59
QY 245 DGLGHGTFFVAGVYASMR--COGFAPDAELHIFRVFTNNQVSYTSMFLDAFYALTKI 301
Db 60 DSSSHGTHVAGTTAALNNSIGVLGVSPASLIVAKYLDSTGSGOYSLINGIEMALSNM 119
QY 302 DVNLISIGDPDMD--HPFVDKYWELTANNVIVSAIGNDP--LYGTLNPPADQMDVIG 357
Db 120 DVINMSLGGPTGSTALLKTYVDKA--VSSGIYVAAAAGNCGSSGSTVGYPAKYPSTIA 176
QY 358 VGGIDEDNINARSSRGMTTWELPGYGRMKPDIVYVYAGVRSYKGGCRALSGTSVAS 417
Db 177 VGVAVNSNORASFSAGS--EL-----DVMAPGVSIQSTLPGITGAYGANGTSMAT 224
QY 418 PVVAGAVTLLVSTVOKRELVPASMKOALIASARLPGVMPFGQHGKLDLLRAYQ 473
Db 225 PHVAGAAALILS--KHPTWTNAQVRDLRESTATYIG--NSFYGGKGLINVOAAQ 275

RESULT 23

US-07-923-260A-3

Sequence 3, Application US/07923260A

Patent No. 5719021

GENERAL INFORMATION:

APPLICANT: Inouye, Masayori

TITLE OF INVENTION: PROTEIN ACTIVATION

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESS: Gerard J. Welser

STREET: 230 South Fifteenth Street, Suite 500

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19102

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/923,260A

FILING DATE: 31-JUL-1992

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Welser, Gerard J.

REGISTRATION NUMBER: 19,763

REFERENCE/DOCKET NUMBER: 377,563BP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-875-8383

TELEFAX: 215-875-8394

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 352 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Bacillus amylosacchariticus

FEATURE:

NAME/KEY: Region

LOCATION: 1..77

OTHER INFORMATION: /note- "The region from 1 to 77 is

OTHER INFORMATION: a propeptide."

US-07-923-260A-3

Query Match

5.8%; Score 323.5; DB 1; Length 352;

	Best local similarity	31.4%;	Pred. No.	5.5e-18;	
	Matches	95;	Conservative	54;	Mismatches 123; Indels 31; Gaps 11.
OY	180	HSRRLLRAIPROVAYOTLQADVLMQMGYGTGANNVAVAFPDGSEKHHPFKVVKERTWMTN	239		
		: : : : : : : : : : : : : : : : : : : : :			
Db	72	HVAHAYKASVPYGVQ- IKAPALHSQGYTGSNNVAVANIDSGIDSHPL- NVKRGASFVP	129		
OY	240	ERT- LDDGCGHGTFAAGVYASMR- --CQSFAPDAELHFFVYFNNQVYSTSFEDAEF	294		
		: : : : : : : : : : : : : : : : : : : : :			
Db	130	SETNPQDQSGSHGHVAGVTAALNNISICVLSVAASALVAKVADLSTSSGQSYIITIGIE	189		
OY	295	YAIKKRIDVNLISIGDFEMD- HPFYDKWELTANNVIMVSAIGNDP- LYETLLNPPA	350		
		: : : : : : : : : : : : : : : : : : : : :			
Db	190	WAISSINDVDYINMISIGBPSGSTALKTVDKA- --VSSGSIYAAAGNESSSSSSSTVQPPA	246		
OY	351	DOMDVIYGVGIDEDINIAFFSSRGRTWELTGGYGRKKPDIYTYGAGVAGRGVGGCCRAL	410		
		: : : : : : : : : : : : : : : : : : : : :			
Db	247	KTPSVIAAGVANNSSNORASFSSVGS- --EL-----DVNAPGYSIOSTLPLPGTYGAX	294		
OY	411	SGTSSVAPVVAAGATLLTYVQKRELIVNAPSMKQALIASARRLPGNNFEDGSHKDLLE	470		
		: : : : : : : : : : : : : : : : : : : : :			
Db	295	NTSMAPFHVAGAAAILTS- --KHPTWTNAAVDRRELESTATYLG- --DSFYYGKGLINVOA	349		
OY	471	AYQ	473		
Db	350	AAQ	352		

```

RESULT 24
US-09-178-155-2
Sequence 2, Application US/09178155
Patent No. 6312936
GENERAL INFORMATION:
APPLICANT: Poulos, Aycockaran J.
APPLICANT: Schellenberger, Volker
APPLICANT: Kellis, Jr. James T.
APPLICANT: Pasch, Christian
APPLICANT: Nadherny, Joanne
APPLICANT: Nakl, Donald P.
TITLE OF INVENTION: Multiply-Substituted Protease Variants
FILE REFERENCE: GC502-2
CURRENT APPLICATION NUMBER: US/09/178,155
CURRENT FILING DATE: 1998-10-23
EARLIER APPLICATION NUMBER: 08/956,323
EARLIER FILING DATE: 1997-10-23
EARLIER APPLICATION NUMBER: 08/956,564
EARLIER FILING DATE: 1997-10-23
EARLIER APPLICATION NUMBER: 08/956,224
EARLIER FILING DATE: 1997-10-23
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO. 2
LENGTH: 382
TYPE: PR1
ORGANISM: B. amylioliquefaciens
US-09-178-155-2

```

Query Match	5.7%	Score 322	DB 4	Length 382
Best Local Similarity	26.2%	Pred. No. 8.2e-18		
Matches 119; Conservative 57; Mismatches 142; Indels 136; Gaps 18				
Oy	1	MLTVNITLLITLVYLLCGKTHLDRLREKKSFEKAPCPGCSHLTLKVEFSSTV----	52	
	1		
Db	1	MGRKRVNLSLFLALA-----LITMVFSTSSAQAAKSN	35	
Oy	53	EYEITVAFNCFYFAKARNSFISALAKSEVDNRRILPRNNSSDPSPDFEYIQIKTEOK	111	
	36	GEKTIIVG-----KPTMTSMNAKKDVI-----SEK	63	
Db	112	AGLITLLE-DHPNIRKVTPOKRVYSIKTAIESDPVPCNSTRMSQKMOSSRPLIRASLSLG	170	
Oy	64	GGRVOKRKYVDASVYLINKEAVTEK--KDPVA--	96	
Db				

Qy	171	SGFWATGHSSRLRLALIPROAQTLOADVLMQWQSYTANRAVAVDGLSKRHHK	429
Db	97	---YEEEDHAAHVAASVPYGVSO- IKAPALSHOGYGTSMNKAAVIDSIGTSHSPDLKV	151
Qy	230	----YKERTNMTNERTLTDDGLGHGTVAQVAVLSRE- --CGFAPDAELIFRFPTNN	281
Db	152	ASGASVSPSETN- --PFQDNNSHGTVAAQTVAAALNNSIGVLGAAPSASLVAVVLGAD	206
Qy	282	QVSTYTFPLDANPFAILKRLIDVNLNLSIGPPEMD- -HPVDEWELNANNVYVLSAIGDN	339
Db	207	GSQGYSMITINGIEMAIANNADVYINMSLGGPSGSAALKAAYDEA- -VAAGVYVVAALAGNE	263
Qy	340	GP--LXGTLANPDADQMDYVGGVIGFEEDINIRFSESGMTWLTLCGGYGRKKPIYVAG	397
Db	264	CTGSSSTYTGPKRTPYVAVGAVDSSNQKASFSTGVP- --EL-----DYMAPGVS	311
Qy	398	VRGS- --GKGGCRALSTVASYVVAAGAVTLVLS	429
Db	312	IQSLTPGNKYG- -AYNQTSMASPTHAAGAAALILIS	343

RESULT 25
5472855-2
Patent No. 5472855
Applicant: CARTER, PAUL J.; WELLS, JAMES A.
TITLE OF INVENTION: SUBSTRATE ASSISTED CATALYSIS
NUMBER OF SEQUENCES: 31
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/287,964
FILING DATE: 22-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 90,902
FILING DATE: 12-JUL-1993
APPLICATION NUMBER: 823,039
FILING DATE: 14-JAN-1992
APPLICATION NUMBER: 35,652
FILING DATE: 06-APR-1987
APPLICATION NUMBER: 334,081
FILING DATE: 04-APR-1989
APPLICATION NUMBER: 127,134
FILING DATE: 01-DEC-1987
APPLICATION NUMBER: 846,627
FILING DATE: 01-APR-1986
APPLICATION NUMBER: 858,594
FILING DATE: 30-APR-1986
APPLICATION NUMBER: 614,612
FILING DATE: 29-MAY-1984
APPLICATION NUMBER: 614,615
FILING DATE: 29-MAY-1984
APPLICATION NUMBER: 614,617
FILING DATE: 29-MAY-1984
APPLICATION NUMBER: 614,491
FILING DATE: 29-MAY-1984
SEQ ID NO: 2;
LENGTH: 382
5472855-2

	Query Match	Best Local Similarity	5.7%;	Score 321.5;	DB 6;	Length 382;
			31.7%;	Pred. No. 9e-18;		
	Matches	98;	Conservative	51;	Mismatches	117;
					Indels	43;
					Gaps	13;
QY	180	HSRRLRLRAIPQVAVQTLQADVLVMQGYGAVNVAVFDTGLSEKHPRFK-----NPK	23;			
DB	102	HVAIAVAGSPVQVSGQ--KKAPALSHSGGYSGVSVAVADISGIDSHPLTKAGAGASMP	160			
QY	234	RTNMTNERTLDGCGHGFPAAGVATIAMRE---CCGFAPDAELHIFRYVTNNQGVSYWPL	290			
DB	161	ETN-----PFDQNNHGHGHAQVAAALNNISIGVLGVAAPSALSYAVKVLGADSGSQY	211			
QY	291	DAEYVALIKKIDYVNLSTSGGPDMD--HEFDKXWELTANNYINSAIGANGP--LYGLT	340			
DB	216	NGEIAALANNMDDYINMSIGGSGSAALKAADYKA---VASGYYVVAALAGNEIGSSSY	277			

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:27:27 ; Search time 75 Seconds
(without alignments)
1869.062 Million cell updates/sec

Title: US-09-830-837-6
Perfect score: 5617
Sequence: 1 MLLVNTWLLVLLGKKH.....PRVKRPLMQVHPPTPSV 1052

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 13350620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

No.	Score	Match	Length	ID	Description
1	5617	100.0	1052	21 AAB06334	Human subtilisin-k
2	5617	100.0	1052	21 AAY84227	Amino acid sequenc
3	5617	100.0	1052	21 AAY66982	Human subtilisin.
4	5617	100.0	1052	21 AAY54619	Human subtilisin p
5	5617	100.0	1052	22 AAY38520	Human subtilisin
6	5617	100.0	1052	23 AAE14527	Human site-1 prote
7	5487	97.7	1052	21 AAY84228	Amino acid sequenc
8	5487	97.7	1052	23 AAE14528	Hamster site-1 pro
9	5476	97.5	1052	21 AAB06335	Rat subtilisin-kex
10	5427	96.6	1052	21 AAB06336	Mouse subtilisin-k

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	5617	100.0	1052	21 AAB06334	Human subtilisin-k
2	5617	100.0	1052	21 AAY84227	Amino acid sequenc
3	5617	100.0	1052	21 AAY66982	Human subtilisin.
4	5617	100.0	1052	21 AAY54619	Human subtilisin p
5	5617	100.0	1052	22 AAY38520	Human subtilisin
6	5617	100.0	1052	23 AAE14527	Human site-1 prote
7	5487	97.7	1052	21 AAY84228	Amino acid sequenc
8	5487	97.7	1052	23 AAE14528	Hamster site-1 pro
9	5476	97.5	1052	21 AAB06335	Rat subtilisin-kex
10	5427	96.6	1052	21 AAB06336	Mouse subtilisin-k

11	3549	63.2	666	23 ABB90255	Human polypeptide
12	2366.5	42.1	992	22 AAB20015	Drosophila site 1
13	2231	39.7	952	22 ABB63047	Drosophila melanog
14	528	9.4	126	22 AAO01619	Human polypeptide
15	340.5	6.1	381	7 AAP60571	Sequence of subtil
16	340.5	6.1	381	9 AAP80744	Sequence of Bacill
17	340.5	6.1	381	11 AAR07970	Subtilisin gene pro
18	338.5	6.0	352	19 AAW46600	Suppressor mutatio
19	337.5	6.0	352	19 AAW46595	Amino acid sequenc
20	337.5	6.0	381	8 AAP71060	Wild-type subtilis
21	337.5	6.0	381	11 AAR03737	Subtilisin gene pr
22	337.5	6.0	381	13 AAR24131	Bacterial serine p
23	337.5	6.0	381	14 AAR34463	Bacillus subtilis
24	337.5	6.0	381	16 AAR74224	B. subtilis subtili
25	337.5	6.0	381	20 AAX39229	Bacillus subtilis
26	336.5	6.0	381	8 AAP70053	Bacillus subtilis
27	336.5	6.0	1079	22 AAB81180	Transglutaminase r
28	335.5	6.0	379	23 AAE19062	Bacillus lichenifo
29	333.5	5.9	275	15 AAR51928	Bacillus subtilis
30	332.5	5.9	381	23 AAE19063	Bacillus subtilis
31	330.5	5.9	275	15 AAR51920	Bacillus subtilis
32	330.5	5.9	275	15 AAR51924	Bacillus subtilis
33	330.5	5.9	275	15 AAR51929	Bacillus subtilis
34	330.5	5.9	275	15 AAR52643	Bacillus subtilis
35	329.5	5.9	275	15 AAR51922	Bacillus subtilis
36	329.5	5.9	275	15 AAR51923	Bacillus subtilis
37	329.5	5.9	275	15 AAR51926	Bacillus subtilis
38	329.5	5.9	275	15 AAR51927	Bacillus subtilis
39	329.5	5.9	379	19 AAW31600	Subtilisin usefu
40	329	5.9	350	19 AAR46598	Amino acid sequenc
41	327.5	5.8	275	12 AAR10207	Mutant subtilisin
42	327.5	5.8	275	15 AAR52644	B. subtilis subtili
43	326.5	5.8	275	12 AAR10206	Mutant subtilisin
44	326.5	5.8	275	12 AAR10208	Mutant subtilisin
45	326.5	5.8	275	12 AAR10205	Mutant subtilisin

ALIGNMENTS

RESULT 1
AAB06334
ID AAB06334 standard; Protein; 1052 AA.
XX AC AAB06334;
XX DT 03-OCT-2000 (first entry)
XX DE Human subtilisin-kexin isoenzyme 1.
XX KW Human; subtilisin-kexin isoenzyme 1; SKI-1; antilipemic; cytostatic;
XX KW vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
XX KW Ras-dependent cancer; restenosis; amyloid protein formation;
XX KW pro-brain-derived neurotrophic factor; proBDNF;
XX KW sterol-regulatory element-binding protein; SREBP.
XX OS Homo sapiens.
XX PN WO200026348-A2.
XX PD 11-MAY-2000.
XX PF 04-NOV-1999; 99WO-CA01058.
XX PR 04-NOV-1998; 98CA-2249548.
XX PA (RECL-) INST RECH CLINIQUES MONTREAL.
XX PI Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
XX DR WPI; 2000-365601/31.
XX DR N-PSDB; AAA57197.

PT Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
 PT producing a polypeptide useful for treating hypercholesterolemia, liver
 PT steatosis and amyloidosis, comprises a specific amino acid sequence -
 XX
 PS
 XX
 Claim 1; Page 56-59; 119pp; English.
 CC The present sequence is human subtilisin-kexin isoenzyme 1 (SKI-1),
 CC a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
 CC from the human neuronal cell line IMR-32 by RT-PCR using active site
 CC degenerate primers. SKI-1 cleaves at a specific threonine residue within
 CC the N-terminal segment of human pro-brain-derived neurotrophic factor
 CC (proBDNF). It is also capable of cleaving a sterol-regulatory
 CC element-binding protein (SREBP), which function to control lipid
 CC biosynthesis and uptake in animal cells. Peptides which bind to and are
 CC treated by SKI-1 may be used for monitoring SKI-1 activity, for screening
 CC inhibitors of SKI-1 activity, or for screening enhancers of SKI-1
 CC catalytic site. Proteic fragments of SKI-1 which bind to the SKI-1
 CC substrate may be used as inhibitors of SKI-1 activity. They may
 CC be used to treat diseases involving overexpression of SKI-1 or SKI-1
 CC substrate. Such diseases include hypercholesterolemia, high levels of
 CC fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,
 CC Ras-dependent cancer, restenosis and amyloid protein formation.
 XX
 SQ Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 21; Length 1052;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLVNIWLLVLLCGKKHGLDRLEKSKFEKAPCGCSHLTLKVFSSVVEYEVAF 60
 DB 1 MKLVNIWLLVLLCGKKHGLDRLEKSKFEKAPCGCSHLTLKVFSSVVEYEVAF 60
 QY 61 NGYFTAKARNSISSALKSSEVDNWRIPRNPSSDPSDFEVIQIKKOKAGLLTLEDH 120
 DB 61 NGYFTAKARNSISSALKSSEVDNWRIPRNPSSDPSDFEVIQIKKOKAGLLTLEDH 120
 QY 121 PNKVTQPKRVSLSKYAESDTPVPCNETRWQSKQSRPLRASLSLGGFWHATGRH 180
 DB 121 PNKVTQPKRVSLSKYAESDTPVPCNETRWQSKQSRPLRASLSLGGFWHATGRH 180
 QY 181 SRRLLRAIPROVAQTLQADVLQMGYTGANRVAVDFGLSEKHPFKNKERTWNINE 240
 DB 181 SRRLLRAIPROVAQTLQADVLQMGYTGANRVAVDFGLSEKHPFKNKERTWNINE 240
 QY 241 RTLDGLGHGTFVAGVIASNRECQGFADAEHLHFRVFTNNQSVYTSWFLDAFNAILKK 300
 DB 241 RTLDGLGHGTFVAGVIASNRECQGFADAEHLHFRVFTNNQSVYTSWFLDAFNAILKK 300
 QY 301 IDVLNLSIGGPDMDHFFVDKVMELTANNVIMVSAIGNDGPLYGLTANNPQDMQDVIYGG 360
 DB 301 IDVLNLSIGGPDMDHFFVDKVMELTANNVIMVSAIGNDGPLYGLTANNPQDMQDVIYGG 360
 QY 361 IDFNEDNIAFSSRGMTTWELPGGYGRMKPDIVTYGAGVSGVKGGRALSGTSVASPVV 420
 DB 361 IDFNEDNIAFSSRGMTTWELPGGYGRMKPDIVTYGAGVSGVKGGRALSGTSVASPVV 420
 QY 421 AGAVTLLYSTVOKRELVPASMKQALIASARPLPGVNMFEQGHGKLDLLRAYOILNSYKP 480
 DB 421 AGAVTLLYSTVOKRELVPASMKQALIASARPLPGVNMFEQGHGKLDLLRAYOILNSYKP 480
 QY 481 QASLSFSDIDTECPYMWPCQPIYGGMTVNVNITLNGMGVYTGRIYDKPQDPYLPQ 540
 DB 481 QASLSFSDIDTECPYMWPCQPIYGGMTVNVNITLNGMGVYTGRIYDKPQDPYLPQ 540
 QY 541 NGDNIEVAFSSVSLWFWGSLAISIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 DB 541 NGDNIEVAFSSVSLWFWGSLAISIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 QY 601 QTSTVKLPKIKYIIPTPPKSRKVLMDQVHNLRYPYPPGFPDRNLKMKNDPLDWDNGDHIHTN 660
 DB 601 QTSTVKLPKIKYIIPTPPKSRKVLMDQVHNLRYPYPPGFPDRNLKMKNDPLDWDNGDHIHTN 660

QY 661 FDMYQHLSRMNGYFVVLGAPPTCFDASQGTGTLTMDSEETFPPEIAKLRRDYDNGLSL 720
 DB 661 FDMYQHLSRMNGYFVVLGAPPTCFDASQGTGTLTMDSEETFPPEIAKLRRDYDNGLSL 720
 QY 721 VIFSDWYNTSVMRKVKFYDENTROWMPDGTGGANIPALNELLSSVNMNGFSDGLYEGEFTL 780
 DB 721 VIFSDWYNTSVMRKVKFYDENTROWMPDGTGGANIPALNELLSSVNMNGFSDGLYEGEFTL 780
 QY 781 ANHDMYASGCSIAKEPPEGVVITOTFKDQGLEVLKQETAVVENVPIILGYOIPAEGGGR 840
 DB 781 ANHDMYASGCSIAKEPPEGVVITOTFKDQGLEVLKQETAVVENVPIILGYOIPAEGGGR 840
 QY 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPERM 900
 DB 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPERM 900
 QY 901 EGNHLHRYSKVLEAHLGDPKPRLPACPRLSWAKPOPLNETAPSNLWKHKLLSIDLDKV 960
 DB 901 EGNHLHRYSKVLEAHLGDPKPRLPACPRLSWAKPOPLNETAPSNLWKHKLLSIDLDKV 960
 QY 961 VLPFRSNRQVRLSPGSGGAWDIPGGMGRYNOVGOTIPVFAFLGAMVLAFFVQ 1020
 DB 961 VLPFRSNRQVRLSPGSGGAWDIPGGMGRYNOVGOTIPVFAFLGAMVLAFFVQ 1020
 QY 1021 INKAKSRPKRRKPRVRKRPOLMQOVHPKTPSV 1052
 DB 1021 INKAKSRPKRRKPRVRKRPOLMQOVHPKTPSV 1052

RESULT 2
 AAY84227
 ID AAY84227 standard; peptide; 1052 AA.
 AC AAY84227;
 XX 03-JUL-2000 (first entry)
 XX Amino acid sequence of a human site-1 protease.
 XX Modulator; sterol-regulated site-1 protease; cholesterol;
 KW sterol regulatory element binding protein; SREBP; lipid synthesis;
 KW fatty acid biosynthesis; site-1 protease; protease inhibitor;
 KW serum cholesterol; hypercholesterolemia; lipid metabolism.
 XX Homo sapiens.
 XX WO200009677-A2.
 XX 24-FEB-2000.
 XX 13-AUG-1999; 99WO-US18544.
 XX 14-AUG-1998; 98US-0096571.
 XX 23-JUL-1999; 99US-0360237.
 XX (TEXA) UNIV TEXAS SYSTEM.
 PA (BROW) BROWN M S.
 PA (CHEN) CHENG D.
 PA (ESPE) ESPENSHADE P J.
 PA (GOLD) GOLDSTEIN J L.
 PA (RAWS) RAWSON R B.
 PA (SAKA) SAKAI J.
 XX Brown MS, Cheng D, Espenshade PJ, Goldstein JL, Rawson RB;
 PI Sakai J;
 XX WPI; 2000-224327/19..
 XX Novel assay for identifying modulators of sterol-regulated Site-1
 PT protease useful for the treatment of hypercholesterolemia, involves
 PT identifying an agent capable of down regulating Site-1 protease
 PT activity

PS Claim 62; Fig 4A-B; 172pp; English.

CC The specification describes a method for identifying modulators of a
CC steroid-regulated site-1 protease. Site-1 protease cleaves sterol
CC regulatory element binding proteins (SREBPs) in the endoplasmic
CC reticulum, initiating release from membranes and activating lipid
CC synthesis. The modulators therefore also modulate cholesterol and
CC fatty acid biosynthesis. The method comprises selecting an agent capable
CC of down regulating site-1 protease and formulating a composition
CC comprising the agent. The site-1 protease inhibitors are useful for
CC treating a patient for elevated serum cholesterol. Diseases treated
CC include hypercholesterolemia and other lipid metabolism associated
CC conditions. The present sequence represents a human site-1 protease.

XX Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 21; Length 1052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLVNWLVLVLLVLLCGKKHGLDRLEKKSFEKAPCGCSHLTLKVFSSVVEYIVAF 60
Db 1 MKLVNWLVLVLLVLLCGKKHGLDRLEKKSFEKAPCGCSHLTLKVFSSVVEYIVAF 60
QY 61 NGYFTAKARNSFTISSALKSSEVDNWRILIPRNPSSDYPDFEVIQIKKQKAGLLLEHDH 120
Db 61 NGYFTAKARNSFTISSALKSSEVDNWRILIPRNPSSDYPDFEVIQIKKQKAGLLLEHDH 120
QY 121 PNIRKRYTPQKRVSRKLYAESDPTVPCNETRWSOKWQSSRPLRRASLSLGSFGFWHATGRH 180
Db 121 PNIRKRYTPQKRVSRKLYAESDPTVPCNETRWSOKWQSSRPLRRASLSLGSFGFWHATGRH 180
QY 181 SSRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVFDTGLSEKHPHFKNVKBERTWNTNE 240
Db 181 SSRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVFDTGLSEKHPHFKNVKBERTWNTNE 240
QY 241 RTLDGGLGHTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
Db 241 RTLDGGLGHTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
QY 301 IDVLNLSIGGPDMPDFVVKWELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360
Db 301 IDVLNLSIGGPDMPDFVVKWELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360
QY 361 IDFDNIARFSSRGMTTWELPGYGRMKPDIVTYGAGVRSYVKGCRALSGTSVSPV 420
Db 361 IDFDNIARFSSRGMTTWELPGYGRMKPDIVTYGAGVRSYVKGCRALSGTSVSPV 420
QY 421 AGAVTLVSTVQKRELNVNPSMKQALIASARLLPGVNMFEQGHGKLDLLRAYQILNSYRP 480
Db 421 AGAVTLVSTVQKRELNVNPSMKQALIASARLLPGVNMFEQGHGKLDLLRAYQILNSYRP 480
QY 481 QASLSPSYIDLTCPYMWPCSOPIYIGMPTVVNTILNGMGTGRIYDKPDQWQYLPQ 540
Db 481 QASLSPSYIDLTCPYMWPCSOPIYIGMPTVVNTILNGMGTGRIYDKPDQWQYLPQ 540
QY 541 NGDNIEVAISYSSVLPWNSGYLAISVTKKAASWEGIAQGHVIMFVSPAETESKNGAE 600
Db 541 NGDNIEVAISYSSVLPWNSGYLAISVTKKAASWEGIAQGHVIMFVSPAETESKNGAE 600
QY 601 QTSTVKLPKIKVPIPTPPRSKRVLDQYHNLRYPCGYFPRDNLKMKNDPLDNGDHIHN 660
Db 601 QTSTVKLPKIKVPIPTPPRSKRVLDQYHNLRYPCGYFPRDNLKMKNDPLDNGDHIHN 660
QY 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGTLLMVDSEEEYFPEIAKLRRDNDGLSL 720
Db 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGTLLMVDSEEEYFPEIAKLRRDNDGLSL 720
QY 721 VIFSDWYNTSVARKVKFYDENTRQWNPDTGGANIPALNELLISVNMWGSGLYEGEFTL 780
Db 721 VIFSDWYNTSVARKVKFYDENTRQWNPDTGGANIPALNELLISVNMWGSGLYEGEFTL 780
QY 781 ANHDMYASGCSIAKFPEDGVVITQTFKQDGLVLEKQETAVVENVPILGYQIPAGGGGR 840

Db 781 ANHDMYASGCSIAKFPEDGVVITQTFKQDGLVLEKQETAVVENVPILGYQIPAGGGGR 840
QY 841 IVLYGDSNCLDDSHROKDCFWLLDALLQYTSYVTPPSLSHSGNRORPPSGAGSVTPERM 900
Db 841 IVLYGDSNCLDDSHROKDCFWLLDALLQYTSYVTPPSLSHSGNRORPPSGAGSVTPERM 900
QY 901 EGNLHRYSKVLEAHILGDKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDKV 960
Db 901 EGNLHRYSKVLEAHILGDKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDKV 960
QY 961 VLPNFSNRQVPRPLSPGSGANDIPGGIMPGRYNOEVGQTIPIVFAFLGAMVLAFFVQ 1020
Db 961 VLPNFSNRQVPRPLSPGSGANDIPGGIMPGRYNOEVGQTIPIVFAFLGAMVLAFFVQ 1020
QY 1021 INKAKSRPRKRKPRVKRPQMLMQOVHPKTPSV 1052
Db 1021 INKAKSRPRKRKPRVKRPQMLMQOVHPKTPSV 1052

RESULT 3
AAY66982

ID AAY66982 standard; Protein; 1052 AA.

XX AAY66982;

DT 15-MAR-2000 (first entry)

DE Human subtilisin.

XX Subtilisin; allergy; human; bacterium; protease; epitope; detergent;
KW cosmetic; textile; pet food industry; debromination treatment.

XX Homo sapiens.

XX WO953078-A2.

XX 21-OCT-1999.

XX 15-APR-1999; 99WO-US08177.

XX 15-APR-1998; 98US-0060854.

XX (GEMV) GENENCOR INT INC.

XX Estell D;

XX WPI; 2000-061971/05.

XX Modified non-human protease having reduced allergenicity -

XX Disclosure; Fig 6; 38pp; English.

XX This sequence represents the human subtilisin protein. The invention
XX relates to a method of reducing the allergenicity of a non-human
XX protein, especially a bacterial protease such as subtilisin (AAY66980),
XX by identifying an epitope on the non-human protein and replacing it by
XX an analogous region from the human protein. The method is useful for
XX producing proteins, including proteases, that are less likely to cause
XX allergic reactions, thus the invention can be employed for example in
XX detergents, cosmetics, textile treatment, and pet food industries. The
XX human subtilisin can be used pharmaceutically for debromination
XX treatments.

XX Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 21; Length 1052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLVNWLVLVLLVLLCGKKHGLDRLEKKSFEKAPCGCSHLTLKVFSSVVEYIVAF 60
Db 1 MKLVNWLVLVLLVLLCGKKHGLDRLEKKSFEKAPCGCSHLTLKVFSSVVEYIVAF 60

Db 241 RTLDGLGHGTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYSYTWFLDAFNALYALKK 300
QY 301 IDVLNLSIGGPDPMDFVVDKWLWELTANNVIMVSAIGNDGPLYGTTLNNDQMDVIGVGG 360
Db 301 IDVLNLSIGGPDPMDFVVDKWLWELTANNVIMVSAIGNDGPLYGTTLNNDQMDVIGVGG 360
QY 361 IDPEDNIARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVGGKGRALSGTSVSPVV 420
Db 361 IDPEDNIARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVGGKGRALSGTSVSPVV 420
QY 421 AGAVTLLVSTVQRELNVNPMASMKOALIASARRLPVGNMFEQGHGKLDLLRAYOILANSYK 480
Db 421 AGAVTLLVSTVQRELNVNPMASMKOALIASARRLPVGNMFEQGHGKLDLLRAYOILANSYK 480
QY 481 QASLSPSYIDLTECPYMWPCSOPIYYGGMPTVVNTILNGMGTGRIVDKDPQWYLPQ 540
Db 481 QASLSPSYIDLTECPYMWPCSOPIYYGGMPTVVNTILNGMGTGRIVDKDPQWYLPQ 540
QY 541 NGDNIEVAFSSVSLVPMWPGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
Db 541 NGDNIEVAFSSVSLVPMWPGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
QY 601 QSTVTKLPKVKLIPTPPSKRVLDQYHNLRYPPGYPFDNLRMKNDPLDWDGDIHTN 660
Db 601 QSTVTKLPKVKLIPTPPSKRVLDQYHNLRYPPGYPFDNLRMKNDPLDWDGDIHTN 660
QY 661 FRDYOHLRSMGYFVEVLGAPTCFDASQVGTLLMVDSEEEYFPEIAKLRRDNDGLSL 720
Db 661 FRDYOHLRSMGYFVEVLGAPTCFDASQVGTLLMVDSEEEYFPEIAKLRRDNDGLSL 720
QY 721 VIFSDWYNTSVMRKVKFYDENTQWMPDGTGGANIPALNELLSSVNMWGFSDGLYEGETL 780
Db 721 VIFSDWYNTSVMRKVKFYDENTQWMPDGTGGANIPALNELLSSVNMWGFSDGLYEGETL 780
QY 781 ANHDMYASGCSIAKPEDGWTOTFKDQGLEVLQKQETAVENVPIGLYQIPABGGGR 840
Db 781 ANHDMYASGCSIAKPEDGWTOTFKDQGLEVLQKQETAVENVPIGLYQIPABGGGR 840
QY 841 IVLYGDSNCLDDSHRQKDFWLLDALLQTSYGTVPSSLSHSGNRRQPPSGAGSVTPERM 900
Db 841 IVLYGDSNCLDDSHRQKDFWLLDALLQTSYGTVPSSLSHSGNRRQPPSGAGSVTPERM 900
QY 901 EGNHLHRYSKVLEAHLGDKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDDKV 960
Db 901 EGNHLHRYSKVLEAHLGDKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDDKV 960
QY 961 VLPNFRSNRQVRLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVVLAFVVQ 1020
Db 961 VLPNFRSNRQVRLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVVLAFVVQ 1020
QY 1021 INKASRPKRKRPRVRKRPQLMQVHPKPTPSV 1052
Db 1021 INKASRPKRKRPRVRKRPQLMQVHPKPTPSV 1052

RESULT 5

AAU38520
ID AAU38520 standard; Protein; 1052 AA.

XX AC AAU38520;

XX DT 18-DEC-2001 (first entry)

XX DE Human subtilisin.

XX KW Immunogenic; immunostimulant; immunosuppressive; T-cell epitope;
KW immune response; hormone; vaccine; cytokine; therapeutic; savinase;
KW hyper-allergenic; hypo-allergenic; cleaning composition; human;
KW textile treatment; contact lens cleaning solution; waste treatment;
KW cosmetic formulation; subtilisin; BPN.
XX OS Homo sapiens.

PN WO200159130-A2.

XX 16-AUG-2001.

XX 22-JAN-2001; 2001WO-US02204.

XX 08-FEB-2000; 2000US-0500135.

XX (GEWV) GENENCOR INT INC.

XX Estell DA, Harding FA;

XX WPI; 2001-607196/69.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

A variant of a polypeptide of interest comprising an altered T-cell epitope to produce a different immunogenic response useful in therapeutics, vaccines, textile treatments and cosmetics -

Example 2; Figure 7; 54pp; English.

The invention relates to a variant of a polypeptide of interest comprising an altered T-cell epitope to produce a different immunogenic response. A method for altering the immunogenicity of a protein especially an enzyme, a hormone, a factor, a vaccine, or cytokine is described. The protein produces an immune response as desired and is useful in therapeutics, vaccines and in forming hyper and hypo-allergenic compounds, e.g. cleaning compositions, textile treatments, contact lens cleaning solutions, waste treatment products and cosmetic formulations. Unlike antibody assays, the method determines immune sensitivity to an epitope prior to any sensitisation taking place. AAU38518-AAU38742 represent the amino acid sequences of Bacillus lentus savinase and human subtilisin (BPN) T-cell epitopes, and related amino acid sequences as described in the method of the invention.

CC Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 22; Length 1052;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLVNVLVLLVLLGKHLGRLEKKSPEKAPCGCSHLTLKVFSSVWEYEVAF 60
Db 1 MKLVNVLVLLVLLGKHLGRLEKKSPEKAPCGCSHLTLKVFSSVWEYEVAF 60
QY 61 NGYFTAKARNSFISSALKSEVDNWRIPRNNPSSDYPSDFEVIQIEKOKAGLLTLEH 120
Db 61 NGYFTAKARNSFISSALKSEVDNWRIPRNNPSSDYPSDFEVIQIEKOKAGLLTLEH 120
QY 121 PNTRKRVTPQRKVFRLKYAESDPTVPCNETRWSQKWSRPLRRASLSLGSFHWATGRH 180
Db 121 PNTRKRVTPQRKVFRLKYAESDPTVPCNETRWSQKWSRPLRRASLSLGSFHWATGRH 180
QY 181 SSRRLRLAIPROVAQTLQADVLWQMGYTGANVRVAVFDTGLSEKHPFKVKTWNTE 240
Db 181 SSRRLRLAIPROVAQTLQADVLWQMGYTGANVRVAVFDTGLSEKHPFKVKTWNTE 240
QY 241 RTLDGGLGHGTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYSYTWFLDAFNALYALKK 300
Db 241 RTLDGGLGHGTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYSYTWFLDAFNALYALKK 300
QY 301 IDVLNLSIGGPDPMDFVVDKWLWELTANNVIMVSAIGNDGPLYGTTLNNDQMDVIGVGG 360
Db 301 IDVLNLSIGGPDPMDFVVDKWLWELTANNVIMVSAIGNDGPLYGTTLNNDQMDVIGVGG 360
QY 361 IDPEDNIARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVGGKGRALSGTSVSPVV 420
Db 361 IDPEDNIARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVGGKGRALSGTSVSPVV 420
QY 421 AGAVTLLVSTVQRELNVNPMASMKOALIASARRLPVGNMFEQGHGKLDLLRAYOILANSYK 480
Db 421 AGAVTLLVSTVQRELNVNPMASMKOALIASARRLPVGNMFEQGHGKLDLLRAYOILANSYK 480
QY 481 QASLSPSYIDLTECPYMWPCSOPIYYGGMPTVVNTILNGMGTGRIVDKDPQWYLPQ 540

Db 481 QASLSPSYDITLTCPTMPCYQPIYYGMPVTNNVILNGMGVTRIVDKPDWQPYLPQ 540
 QY 541 NGDNIEVAFSYSSVLPWPSGVLAISSIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 Db 541 NGDNIEVAFSYSSVLPWPSGVLAISSIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 QY 601 QTSVTKLPKIKVITPTPRSKRVLDQYHNLRYPCYPPRONLRKNDPLDNGDHIHTN 660
 Db 601 QTSVTKLPKIKVITPTPRSKRVLDQYHNLRYPCYPPRONLRKNDPLDNGDHIHTN 660
 QY 661 FDMYQHLRSMGVFEVLGAPFTCFDASQYGTLLMDVSEEEFFEEIAKLKRDVDNGLSL 720
 Db 661 FDMYQHLRSMGVFEVLGAPFTCFDASQYGTLLMDVSEEEFFEEIAKLKRDVDNGLSL 720
 QY 721 VIFSDWNTSVMRKVFYDENTROWMPDGTGGANIPALNELLSSVWNGFSDGLYGEFTL 780
 Db 721 VIFSDWNTSVMRKVFYDENTROWMPDGTGGANIPALNELLSSVWNGFSDGLYGEFTL 780
 QY 781 ANHDMYASGCSIAKPEPVGVIQTFKQDGLVLEKQETAVENVPILGLYQIPAEGGGR 840
 Db 781 ANHDMYASGCSIAKPEPVGVIQTFKQDGLVLEKQETAVENVPILGLYQIPAEGGGR 840
 QY 841 IVLYGDSNCLDDSHRQKDCFLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
 Db 841 IVLYGDSNCLDDSHRQKDCFLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
 QY 901 EGNHLHRYSKVLEAHLGDKPRPLPACPRLSWAKPQPLNETAPSNLWKHOKLISIDLQV 960
 Db 901 EGNHLHRYSKVLEAHLGDKPRPLPACPRLSWAKPQPLNETAPSNLWKHOKLISIDLQV 960
 QY 961 VLPNFRNRQVRPLSPGSGAWDIPGGIMPGRYNOEYVGTIPVFAFLGAMVYLAFFVYQ 1020
 Db 961 VLPNFRNRQVRPLSPGSGAWDIPGGIMPGRYNOEYVGTIPVFAFLGAMVYLAFFVYQ 1020
 QY 1021 INKASRPRKRRKVRKRPQMQVHPKPTSV 1052
 Db 1021 INKASRPRKRRKVRKRPQMQVHPKPTSV 1052

RESULT 6

AE14527
 ID AE14527 standard; Protein; 1052 AA.
 AC AE14527;
 XX
 DT 07-MAY-2002 (first entry)
 XX
 DE Human site-1 protease.
 XX
 KW Site-1 protease; SLP; promoter; transcription; obesity; type II diabetes;
 KW sterol regulatory element-binding protein; SREBP; hypercholesterolemia;
 KW dyslipidaemia; atherosclerosis; cardiovascular disease; human.
 XX
 OS Homo sapiens.
 XX
 PN W0200200873-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 19-JUN-2001; 2001WO-SE01386.
 XX
 PR 27-JUN-2000; 2000SE-0002417.
 XX
 PA (BIOV-) BIOVITRUM AB.
 XX
 PI Abrahamson L, Ekblom J, Forsgren M, Hoerling J, Johansson P;
 XX
 DR WPI; 2002-139918/18.
 XX
 DR N-PSDB; A024182.
 XX
 PT Human site-1 protease promoter region for identifying agents capable of
 inhibiting the promoter activity useful in treating medical conditions

PT XX such as obesity, diabetes, atherosclerosis and hypercholesterolemia
 PS Disclosure; Page 22-24; 36pp; English.
 CC The invention relates to human site-1 protease (SLP) promoter region.
 CC The promoter sequence is useful for identification of compounds that
 CC inhibit transcription of SLP which in turn results in inhibition
 CC of sterol regulatory element-binding protein (SREBP) pathway. The
 CC compound identified is useful for the treatment of medical conditions
 CC related to obesity, type II diabetes, hypercholesterolemia,
 CC dyslipidaemia, atherosclerosis and other cardiovascular diseases.
 CC The present sequence is human site-1 protease.
 XX
 XX Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 23; Length 1052;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLVNIMLLLVLLCGKHLGDRLEKSKFEKAPCCPSHLTKVFSSVVEYIYAF 60
 Db 1 MKLVNIMLLLVLLCGKHLGDRLEKSKFEKAPCCPSHLTKVFSSVVEYIYAF 60
 QY 61 NGYFTAKARNSFISSALKSEVDNRIIPRNPSSDYPDFEVIQIKERKAGLLTLEH 120
 Db 61 NGYFTAKARNSFISSALKSEVDNRIIPRNPSSDYPDFEVIQIKERKAGLLTLEH 120
 QY 121 PNIRKVTTPQRKRVFSLKYAESDPTVPCNETRWSOKWSSRPLRRASLSLGSFWHATGRH 180
 Db 121 PNIRKVTTPQRKRVFSLKYAESDPTVPCNETRWSOKWSSRPLRRASLSLGSFWHATGRH 180
 QY 181 SSRLLRAIPROVAQTLQADVLWQGYTGANRVAVFDTGLSEKHPHFKNKERTNWTNE 240
 Db 181 SSRLLRAIPROVAQTLQADVLWQGYTGANRVAVFDTGLSEKHPHFKNKERTNWTNE 240
 QY 241 RTLDGLGHGTFVAGYVASMRCQGFAPDAELHIFRVFTNNQVSYTSFLEDAFYALKK 300
 Db 241 RTLDGLGHGTFVAGYVASMRCQGFAPDAELHIFRVFTNNQVSYTSFLEDAFYALKK 300
 QY 301 IDVNLSTGGPDMHDPFVDKWTANNVWISAIGNDGLYGLTNLNPADOMDVIYGGV 360
 Db 301 IDVNLSTGGPDMHDPFVDKWTANNVWISAIGNDGLYGLTNLNPADOMDVIYGGV 360
 QY 361 IDFNITARSSRGMTTWELPGYGRMKPDIYTGAGVRGSGVKGCRALSGTSVASPVV 420
 Db 361 IDFNITARSSRGMTTWELPGYGRMKPDIYTGAGVRGSGVKGCRALSGTSVASPVV 420
 QY 421 AGAVTLVSTVQKRELVPASMKQALIASARRLPVGNPQGHGKLDLLRAYQILNSTKP 480
 Db 421 AGAVTLVSTVQKRELVPASMKQALIASARRLPVGNPQGHGKLDLLRAYQILNSTKP 480
 QY 481 QASLSPSYDITLTCPTMPCYQPIYYGMPVTNNVILNGMGVTRIVDKPDWQPYLPQ 540
 Db 481 QASLSPSYDITLTCPTMPCYQPIYYGMPVTNNVILNGMGVTRIVDKPDWQPYLPQ 540
 QY 541 NGDNIEVAFSYSSVLPWPSGVLAISSIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 Db 541 NGDNIEVAFSYSSVLPWPSGVLAISSIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 QY 601 QTSVTKLPKIKVITPTPRSKRVLDQYHNLRYPCYPPRONLRKNDPLDNGDHIHTN 660
 Db 601 QTSVTKLPKIKVITPTPRSKRVLDQYHNLRYPCYPPRONLRKNDPLDNGDHIHTN 660
 QY 661 FDMYQHLRSMGVFEVLGAPFTCFDASQYGTLLMDVSEEEFFEEIAKLKRDVDNGLSL 720
 Db 661 FDMYQHLRSMGVFEVLGAPFTCFDASQYGTLLMDVSEEEFFEEIAKLKRDVDNGLSL 720
 QY 721 VIFSDWNTSVMRKVFYDENTROWMPDGTGGANIPALNELLSSVWNGFSDGLYGEFTL 780
 Db 721 VIFSDWNTSVMRKVFYDENTROWMPDGTGGANIPALNELLSSVWNGFSDGLYGEFTL 780
 QY 781 ANHDMYASGCSIAKPEPVGVIQTFKQDGLVLEKQETAVENVPILGLYQIPAEGGGR 840

Db 781 ANHDMYASGCSIAKPEDGVVITQTFKQGLEVLKQETAVENVPIGLYQIYPAEGGR 840
QY 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900
Db 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900
QY 901 EGNHLHRYSKVLEAHILGDKPRPLPACPRLSWAKPQPLNETAPSNLWKHKLLSIDLDKV 960
Db 901 EGNHLHRYSKVLEAHILGDKPRPLPACPRLSWAKPQPLNETAPSNLWKHKLLSIDLDKV 960
QY 961 VLPNFRSNRPOVRPLSPGSGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVVLAFVQ 1020
Db 961 VLPNFRSNRPOVRPLSPGSGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVVLAFVQ 1020
QY 1021 INKAKSRPRKRPRVRPQLMQQVHPKTPSV 1052
Db 1021 INKAKSRPRKRPRVRPQLMQQVHPKTPSV 1052

RESULT 7
ID AAY84228
AC AAY84228
XX AAY84228
DT 03-JUL-2000 (first entry)
XX Amino acid sequence of a hamster site-1 protease.
DE Modulator; sterol-regulated Site-1 protease; cholesterol;
KW sterol regulatory element binding protein; SREBP; lipid synthesis;
KW fatty acid biosynthesis; site-1 protease; protease inhibitor;
KW serum cholesterol; hypercholesterolemia; lipid metabolism.
XX Cricetulus sp.
OS
XX
XX W0200009677-A2.
XX
PD 24-FEB-2000.
XX
PF 13-AUG-1999; 99WO-US18544.
XX
PR 14-AUG-1998; 98US-0096571.
PR 23-JUL-1999; 99US-0360237.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
PA (BROW/) BROWN M S.
PA (CHEN/) CHENG D.
PA (ESPE/) ESPENSHADE P J.
PA (GOLD/) GOLDSTEIN J L.
PA (RAWS/) RAWSON R B.
PA (SAKA/) SAKAI J.
XX
PI Brown MS, Cheng D, Espenshade PJ, Goldstein JL, Rawson RB;
PI Sakai J;
XX
DR WPI: 2000-224327/19.
XX
PT Novel assay for identifying modulators of sterol-regulated Site-1
PT protease useful for the treatment of hypercholesterolemia, involves
PT identifying an agent capable of down regulating Site-1 protease
PT activity
XX
XX Claim 60; Fig 4A-B; 172pp; English.
XX
CC The specification describes a method for identifying modulators of a
CC sterol-regulated Site-1 protease. Site-1 protease cleaves sterol
CC regulatory element binding proteins (SREBPs) in the endoplasmic
CC reticulum, initiating release from membranes and activating lipid
CC synthesis. The modulators therefore also modulate cholesterol and
CC fatty acid biosynthesis. The method comprises selecting an agent capable
CC of down regulating Site-1 protease and formulating a composition
CC comprising the agent. The site-1 protease inhibitors are useful for

CC treating a patient for elevated serum cholesterol. Diseases treated
CC include hypercholesterolemia and other lipid metabolism associated
CC conditions. The present sequence represents a hamster site-1 protease.
XX
SQ Sequence 1052 AA;
Query Match 97.7%; Score 5487; DB 21; Length 1052;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;
QY 1 MKLVNWLVLVLLCGKHLGDRLEKSKFEKAPCCSHLTLLKVFSSSTVVEYEVAF 60
Db 1 MKLVNWLVLVLLCGKHLGDRLEKSKFEKAPCCSHLTLLKVFSSSTVVEYEVAF 60
QY 61 NGYFTAKARNSFISSALKSESDVNRWIIIPRNNSDDYPDFEVIQIEKQKAGLLTLEDH 120
Db 61 NGYFTAKARNSFISSALKSESDVNRWIIIPRNNSDDYPDFEVIQIEKQKAGLLTLEDH 120
QY 121 PNTRVTPQRKVRSLKYAESDPTVPCNETRWQKQSSRPLRRASLSLGSFGFWHATGRH 180
Db 121 PNTRVTPQRKVRSLKYAESDPTVPCNETRWQKQSSRPLRRASLSLGSFGFWHATGRH 180
QY 181 SSRLLRAIPROVAQTLQADVLQMGYTGANVRVAVEDTGLSEKHPFKVKTERTWNE 240
Db 181 SSRLLRAIPROVAQTLQADVLQMGYTGANVRVAVEDTGLSEKHPFKVKTERTWNE 240
QY 241 RTLDDGLGHTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYTSWFLDAFYALKK 300
Db 241 RTLDDGLGHTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYTSWFLDAFYALKK 300
QY 301 IDVLNLSIGGPDFMDHPFVKVWELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360
Db 301 IDVLNLSIGGPDFMDHPFVKVWELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360
QY 361 IDFEDNIARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSYKGGCRALSGTSVSPVV 420
Db 361 IDFEDNIARSSRGMTTWELPGGYGRMKPDIVTYGAGVRSYKGGCRALSGTSVSPVV 420
QY 421 AGAVTLVSTVQKRELNVNPMASMKOALIASARRLPGVNMFEQHGKLDLLRAYQILNSYKP 480
Db 421 AGAVTLVSTVQKRELNVNPMASMKOALIASARRLPGVNMFEQHGKLDLLRAYQILNSYKP 480
QY 481 QASLSPSYIDLTECPYMWPCSQPIYYGMPVTVNVTILNGMGVYTGTRVDKQWQPLPQ 540
Db 481 QASLSPSYIDLTECPYMWPCSQPIYYGMPVTVNVTILNGMGVYTGTRVDKQWQPLPQ 540
QY 541 NGDNIEVAFSYSSVLPWPSGYLAISVTKKAASWEGIAQGHVMTITVASPAETESKNGAE 600
Db 541 NGDNIEVAFSYSSVLPWPSGYLAISVTKKAASWEGIAQGHVMTITVASPAETESKNGAE 600
QY 601 QTSVTKLPKVKIIPTPPRSKRVLDQYHNLRYPPGYFPRDNLRMKNDPLDWDGHTN 660
Db 601 HTSTVKLPKVKIIPTPPRSKRVLDQYHNLRYPPGYFPRDNLRMKNDPLDWDGHTN 660
QY 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGTLLMVDSEEEYFPEETAKLRDNDGLSL 720
Db 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGTLLMVDSEEEYFPEETAKLRDNDGLSL 720
QY 721 VIFSDWNTSVMRKVFYDENTROWMPDTGGANIPALNELLSSWNWGFSDGLYEGETL 780
Db 721 VIFSDWNTSVMRKVFYDENTROWMPDTGGANIPALNELLSSWNWGFSDGLYEGETL 780
QY 781 ANHDMYASGCSIAKPEDGVVITQTFKQGLEVLKQETAVENVPIGLYQIYPAEGGR 840
Db 781 ANHDMYASGCSIAKPEDGVVITQTFKQGLEVLKQETAVENVPIGLYQIYPAEGGR 840
QY 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900
Db 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900
QY 901 EGNHLHRYSKVLEAHILGDKPRPLPACPRLSWAKPQPLNETAPSNLWKHKLLSIDLDKV 960
Db 901 EGNHLHRYSKVLEAHILGDKPRPLPACPRLSWAKPQPLNETAPSNLWKHKLLSIDLDKV 960

QY 961 VLPNRSRNPQVRLPGSGGAWDIPGIMPGRYNOEVCOTIPVFAFLGAMVLAFFVQ 1020
 DB 961 VLPNRSRNPQVRLPGSGGAWDIPGIMPGRYNOEVCOTIPVFAFLGAMVLAFFVQ 1020
 QY 1021 INKAKSRPKRRKPRVRRPQMQVHPKTPSV 1052
 DB 1021 ISKAKSRPKRRRPRAKRPOLTOOTHPPTPSV 1052

RESULT 8
 ID AAE14528 standard; Protein; 1052 AA.
 AC AAE14528;
 XX 07-MAY-2002 (first entry)
 DT Hamster site-1 protease.
 DE Site-1 protease; SLP; promoter; transcription; obesity; type II diabetes;
 KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;
 KW dyslipidaemia; atherosclerosis; cardiovascular disease; hamster.
 XX OS Cricetulus griseus.
 XX PN W0200200873-AA.
 XX PD 03-JAN-2002.
 XX PF 19-JUN-2001; 2001WO-S001386.
 XX PR 27-JUN-2000; 2000SE-0002417.
 XX PA (BIOV-) BIOVITRUM AB.
 XX PI Abrahamson L, Ekblom J, Forsgren M, Hoerling J, Johansson P;
 DR WPI; 2002-139918/18.
 DR N-PSDB; AAD24185.
 XX Human site-1 protease promoter region for identifying agents capable of
 PT inhibiting the promoter activity useful in treating medical conditions
 PT such as obesity, diabetes, atherosclerosis and hypercholesterolemia -
 XX PS Disclosure; Page 29-32; 36pp; English.
 CC The invention relates to human site-1 protease (SLP) promoter region.
 CC The promoter sequence is useful for identification of compounds that
 CC inhibit transcription of SLP, which in turn results in inhibition
 CC of sterol regulatory element-binding protein (SREBP) pathway. The
 CC compound identified is useful for the treatment of medical conditions
 CC related to obesity, type II diabetes, hypercholesterolaemia,
 CC dyslipidaemia, atherosclerosis and other cardiovascular diseases.
 CC The present sequence is hamster site-1 protease.
 XX SQ Sequence 1052 AA;

Query Match 97.7%; Score 5487; DB 23; Length 1052;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKLVNIWLLVLLCGKKHLGDRLEKKEKAPCPGCSHLTLKVEFSVVEYEVAF 60
 DB 1 MKLVNIWLLVLLCGKKHLGDRLEKKEKAPCPGCSHLTLKVEFSVVEYEVAF 60
 QY 61 NGYFTAKARNSFTSALKSSEVDNWRILPRNPSDDPSDFEVIQKEKAGLLTLEDH 120
 DB 61 NGYFTAKARNSFTSALKSSEVDNWRILPRNPSDDPSDFEVIQKEKAGLLTLEDH 120
 QY 121 PNKRVTPQKRVSLKVAESDPTVPCNETRWOSKWRPLRRASISLGSCFWHATGRH 180
 DB 121 PNKRVTPQKRVSLKVAESDPTVPCNETRWOSKWRPLRRASISLGSCFWHATGRH 180

QY 181 SSRLLRAIPROVAQTLOADVLMQGYTCANRVAVFDTGLSEKHHPKVKERTWTNE 240
 DB 181 SSRLLRAIPROVAQTLOADVLMQGYTCANRVAVFDTGLSEKHHPKVKERTWTNE 240
 QY 241 RTLDGLGHGTFVAGVIASNRECQGFADAEHLHFRVFTNNQVSYTSWFLDAFNAILKK 300
 DB 241 RTLDGLGHGTFVAGVIASNRECQGFADAEHLHFRVFTNNQVSYTSWFLDAFNAILKK 300
 QY 301 IDVLNLSIGGPDFMDHPFDVKWELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360
 DB 301 IDVLNLSIGGPDFMDHPFDVKWELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360
 QY 361 IDFDNIAFSSRGMTTWELPGYGRMKPDIIVTGAAGVRSYGKGCALSGTSVASPVV 420
 DB 361 IDFDNIAFSSRGMTTWELPGYGRMKPDIIVTGAAGVRSYGKGCALSGTSVASPVV 420
 QY 421 AGAVTLLVSTVOKRELINPASMKQALIASARRLPGVNMFQGHGKLDLLRAYOILNSYK 480
 DB 421 AGAVTLLVSTVOKRELINPASMKQALIASARRLPGVNMFQGHGKLDLLRAYOILNSYK 480
 QY 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMPTVVNVTILNGMGVTCGRIVDKPDQYLPQ 540
 DB 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMPTVVNVTILNGMGVTCGRIVDKPDQYLPQ 540
 QY 541 NGDNIEVAFSSVLPWPGSYLAISVTKKAASWEGIAOCHVIMIVASPAETESKNGAE 600
 DB 541 NGDNIEVAFSSVLPWPGSYLAISVTKKAASWEGIAOCHVIMIVASPAETESKNGAE 600
 QY 601 QTSVTKLPKVKIIPPPRSKRVLMQVHNLRYPGYPFPRDNLRMKNPDLWNGDHIHTN 660
 DB 601 HTSTVTKLPKVKIIPPPRSKRVLMQVHNLRYPGYPFPRDNLRMKNPDLWNGDHIHTN 660
 QY 661 FRDMYQHLRSMGYFEVLGAPTCFDSQYGTLLMVDSEEEYPPEIAKLRRDNDGLSL 720
 DB 661 FRDMYQHLRSMGYFEVLGAPTCFDSQYGTLLMVDSEEEYPPEIAKLRRDNDGLSL 720
 QY 721 VIFSDWYNTSVMRKVFYDENTROWMPDTGGANIPALNELLSSVNMNMGFSDGLYEGEFTL 780
 DB 721 VIFSDWYNTSVMRKVFYDENTROWMPDTGGANIPALNELLSSVNMNMGFSDGLYEGEFTL 780
 QY 781 ANHDMYASGCSIAKPFDEGVVITQFDQGLVLEVKQETAVVENVPILGYQIPAEAGGR 840
 DB 781 ANHDMYASGCSIAKPFDEGVVITQFDQGLVLEVKQETAVVENVPILGYQIPAEAGGR 840
 QY 841 IVLYGDSNCLDSDSHRQKDCFWLLDALLQYTSYGVNPPPSLSHSGNRQRPSPGAGSVTPERM 900
 DB 841 IVLYGDSNCLDSDSHRQKDCFWLLDALLQYTSYGVNPPPSLSHSGNRQRPSPGAGSVTPERM 900
 QY 901 EGNHLHRYSKVLEAHLGDPKPRLPACPLSWAKPQPLNETAPSNLWKHQLLSIDLKDV 960
 DB 901 EGNHLHRYSKVLEAHLGDPKPRLPACPLSWAKPQPLNETAPSNLWKHQLLSIDLKDV 960
 QY 961 VLPNRSRNPQVRLPGSGGAWDIPGIMPGRYNOEVCOTIPVFAFLGAMVLAFFVQ 1020
 DB 961 VLPNRSRNPQVRLPGSGGAWDIPGIMPGRYNOEVCOTIPVFAFLGAMVLAFFVQ 1020
 QY 1021 INKAKSRPKRRKPRVRRPQMQVHPKTPSV 1052
 DB 1021 ISKAKSRPKRRRPRAKRPOLTOOTHPPTPSV 1052

RESULT 9
 ID AAB06335 standard; Protein; 1052 AA.
 AC AAB06335;
 XX 03-OCT-2000 (first entry)
 DT Rat subtilisin-kexin isoenzyme 1.
 DE Rat; subtilisin-kexin isoenzyme 1; SKI-1; antilipase; cytosolic;
 KW Rat; subtilisin-kexin isoenzyme 1; SKI-1; antilipase; cytosolic;

vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
Ras-dependent cancer; restenosis; amyloid protein formation;
pro-brain-derived neurotrophic factor; proBDNF;
sterol-regulatory element-binding protein; SREBP.
Rattus sp.
WO200026348-A2.
11-MAY-2000.
04-NOV-1999; 99WO-CA01058.
04-NOV-1998; 98CA-2249648.
(RECL-) INST RECH CLINIQUES MONTREAL.
Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
WPI: 2000-365601/31.
N-PSDB; AAA571198.
Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
producing a polypeptide useful for treating hypercholesterolemia, liver
steatosis and amyloidosis, comprises a specific amino acid sequence -
Claim 1; Page 61-65; 119pp; English.
The present sequence is rat subtilisin-kexin isoenzyme 1 (SKI-1),
a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
from rat adrenal glands by RT-PCR using active site degenerate
primers. SKI-1 cleaves at a specific threonine residue within
the N-terminal segment of pro-brain-derived neurotrophic factor
(proBDNF). It is also capable of cleaving sterol-regulatory
element-binding proteins (SREBPs), which function to control lipid
biosynthesis and uptake in animal cells. Peptides which bind to and are
cleaved by SKI-1 may be used for monitoring SKI-1 activity, for screening
inhibitors of SKI-1 activity, or for screening enhancers of SKI-1
activity. Proteic fragments of SKI-1 which bind to the SKI-1
catalytic site may be used as inhibitors of SKI-1 activity. They may
be used to treat diseases involving overexpression of SKI-1 or SKI-1
substrate. Such diseases include hypercholesterolaemia, high levels of
fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,
Ras-dependent cancer, restenosis and amyloid protein formation.
SQ Sequence 1052 AA;
Query Match 97.5%; Score 5476; DB 21; Length 1052;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1017; Conservative 20; Mismatches 15; Indels 0; Gaps 0;
QY 1 MKLVNIWLLVLLCGKHLGDRLEKKSPKAPCPGCSHLTLKVFSSVTVEYIVAF 60
DB 1 MKLVNIWLLVLLCGKHLGDRLEKKAPKAPCSHLTLKVFSSVTVEYIVAF 60
QY 61 NGYFTAKARNSFISALKSEVDNRIIPRNPSSDYPSPDFEVIQKEKAGLLTLEDH 120
DB 61 NGYFTAKARNSFISALKSEVDNRIIPRNPSSDYPSPDFEVIQKEKAGLLTLEDH 120
QY 121 PNIRKVTPOKVRPSLKAEASDTPVPCNETRWSQKWSRPLRRASLSLGSFGFWHATGRH 180
DB 121 PNIRKVTPOKVRPSLKAEASDTPVPCNETRWSQKWSRPLRRASLSLGSFGFWHATGRH 180
QY 181 SSRRLRLRAIPROVAOTLQADVLWQMGYTGANVRVAFDTCLEKHPHFKNVKERTWNTNE 240
DB 181 SSRRLRLRAIPROVAOTLQADVLWQMGYTGANVRVAFDTCLEKHPHFKNVKERTWNTNE 240
QY 241 RTLDDGLGHGTFVAGVIASMRQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
DB 241 RTLDDGLGHGTFVAGVIASMRQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
QY 301 IDVLNLSIGGPDFMDHFFVDKVMELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360
:|||||

Db 301 MDVLNLSIGGPDFMDHFFVDKVMELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360
QY 361 IDFDNLTARSSRGMTTWELPGYGRMKPDIIVTYGAGVRGSGVKGCRALSGTSVASPVV 420
Db 361 IDFDNLTARSSRGMTTWELPGYGRMKPDIIVTYGAGVRGSGVKGCRALSGTSVASPVV 420
QY 421 AGAVTLIVSTVQKRELNPNASMKQALIASARRLPVGNMFEGQHGKLDLLRAYQILSYKP 480
Db 421 AGAVTLIVSTVQKRELNPNASMKQALIASARRLPVGNMFEGQHGKLDLLRAYQILSYKP 480
QY 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMTVVNVTILNMGVGTGRIVDKPWQYLPQ 540
Db 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMTVVNVTILNMGVGTGRIVDKPWQYLPQ 540
QY 541 NGDNIEVAFSYSSVLNFWPSGYLAISISVTKKAASWEGIAOGHVMITVASPAETESKNGAE 600
Db 541 NGDNIEVAFSYSSVLNFWPSGYLAISISVTKKAASWEGIAOGHVMITVASPAETESKNGAE 600
QY 601 QTSVTKLPIKVKIIPTPPSKRVLMDQYHNLRYPPGYFPRDNLNRMKNDPLDWDNGDHHTN 660
Db 601 HTSVTKLPIKVKIIPTPPSKRVLMDQYHNLRYPPGYFPRDNLNRMKNDPLDWDNGDHHTN 660
QY 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGTLLMVDSEEEYFPEETAKLRDNDGLSL 720
Db 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGTLLMVDSEEEYFPEETAKLRDNDGLSL 720
QY 721 VIFSDWYNTSVMRKVFYDENTQWMPDTGGANIPALNELLSSVNNMFSGLYEGEFTL 780
Db 721 VIFSDWYNTSVMRKVFYDENTQWMPDTGGANIPALNELLSSVNNMFSGLYEGEFTL 780
QY 781 ANHDMYASGCSIAKFPEDGCVITQTFKDGLEVLKQETAVENVPILGLYQIPAEGGGR 840
Db 781 ANHDMYASGCSIAKFPEDGCVITQTFKDGLEVLKQETAVENVPILGLYQIPAEGGGR 840
QY 841 IVLYGSDNCLDDSHRQKDCFWLLDALLOQTSYGVTPPSLSHSGNRQRPSPGAGSVTTERM 900
Db 841 IVLYGSDNCLDDSHRQKDCFWLLDALLOQTSYGVTPPSLSHSGNRQRPSPGAGSVTTERM 900
QY 901 EGNHLHRSKVLBAHLGDPKPRLPACPRLSWAKPQPLNETAPSNLWKHKLISIDLDKV 960
Db 901 EGNHLHRSKVLBAHLGDPKPRLPACPRLSWAKPQPLNETAPSNLWKHKLISIDLDKV 960
QY 961 VLPNFRSNRPQVRPLSPGSGAWDIPGGIMPGRYNQEVGQTIPIVFAFLGAMVLAFFVQ 1020
Db 961 VLPNFRSNRPQVRPLSPGSGAWDIPGGIMPGRYNQEVGQTIPIVFAFLGAMVLAFFVQ 1020
QY 1021 INKASRPKRKRPRVRKRPQMQVHPKTPSV 1052
Db 1021 ISKASRPKRKRPRVRKRPQMQVHPKTPSV 1052
RESULT 10
AAB06336
ID AAB06336 standard; Protein; 1052 AA.
xx AAB06336;
xx AC AC
xx 03-OCT-2000 (first entry)
xx Mouse subtilisin-kexin isoenzyme 1.
xx Mouse; subtilisin-kexin isoenzyme 1; SKI-1; antilipaeamic; cytostatic;
xx vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
xx Ras-dependent cancer; restenosis; amyloid protein formation;
xx pro-brain-derived neurotrophic factor; proBDNF;
xx sterol-regulatory element-binding protein; SREBP.
xx Mus sp.
xx WO200026348-A2.
xx 11-MAY-2000.
xx

PF 04-NOV-1999; 99WO-CA01058.
PR 04-NOV-1998; 98CA-2249648.
XX (RECL-) INST RECH CLINIQUES MONTREAL.
PA Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
PI WPI; 2000-365601/31.
XX DR N-PSDB; AAA57199.
XX Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
PT producing a polypeptide useful for treating hypercholesterolemia, liver
PT steatosis and amyloidosis, comprises a specific amino acid sequence -
XX
PS Claim 1; Page 67-71; 119pp; English.
XX The present sequence is mouse subtilisin-kexin isoenzyme 1 (SKI-1),
CC a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
CC from mouse corticotropic cells by RT-PCR using active site degenerate
CC primers. SKI-1 cleaves at a specific threonine residue within
CC the N-terminal segment of pro-brain-derived neurotrophic factor
CC (proBDNF). It is also capable of cleaving sterol-regulatory
CC element-binding proteins (SREBPs), which function to control lipid
CC biosynthesis and uptake in animal cells. Peptides which bind to and are
CC cleaved by SKI-1 may be used for monitoring SKI-1 activity for screening
CC inhibitors of SKI-1 activity, or for screening enhancers of SKI-1
CC activity. Protein fragments of SKI-1 which bind to the SKI-1
CC catalytic site may be used as inhibitors of SKI-1 activity. They may
CC be used to treat diseases involving overexpression of SKI-1 or SKI-1
CC substrate. Such diseases include hypercholesterolemia, high levels of
CC fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,
CC Ras-dependent cancer, restenosis and amyloid protein formation.
XX
SQ Sequence 1052 AA;

Query Match 96.6%; Score 5427; DB 21; Length 1052;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 1008; Conservative 25; Mismatches 19; Indels 0; Gaps 0;

QY 1 MKLVIMILLVLLVLLCGKKHLDREKFKAPCPGCSHLTLKVEFSSTVVEYIVAF 60
DB 1 MKLVSTWLLVLLVLLCGKKHLDREKFKAPCPGCSHLTLKVEFSSTVVEYIVAF 60
QY 61 NGYFTAKARNSFSSALKSSEVDNRIIPRNPSSDPSDFEVIQIEKOKAGLLTLEDH 120
DB 61 NGYFTAKARNSFSSALKSSEVDNRIIPRNPSSDPSDFEVIQIEKOKAGLLTLEDH 120
QY 121 PNKEVTPQKVRSLKYAESDPTVPCNETMSOKWSSRPLRSLGSGFWHATGRH 180
DB 121 PNKVVTPQKVRSLKYAESDPTVPCNETMSOKWSSRPLRSLGSGFWHATGRH 180
QY 181 SSRLLRAIPRVAQTLQADVLWQMGYTGANVRVAVFDGLSEKHPFKRVKERTNWTNE 240
DB 181 SSRLLRAIPRVAQTLQADVLWQMGYTGANVRVAVFDGLSEKHPFKRVKERTNWTNE 240
QY 241 RTLDGLGHGTGVAGVIASMRGCGFAPDAELHIFRVFTNNQVSYTSWFLDAFNVAALKK 300
DB 241 RTLDGLGHGTGVAGVIASMRGCGFAPDAELHIFRVFTNNQVSYTSWFLDAFNVAALKK 300
QY 301 IDVLNLSTGGDFMDHPFVDFKVMELTANNVIMVSAIGNDGPLYGLTNLPADQMDVIGVGG 360
DB 301 MDVLNLSTGGDFMDHPFVDFKVMELTANNVIMVSAIGNDGPLYGLTNLPADQMDVIGVGG 360
QY 361 IDFDENIARSSRGMTTWELPGGYGRMKPDIIVTYGAGVRGSGVKGCGCRALSGTSVASPV 420
DB 361 IDFDENIARSSRGMTTWELPGGYGRMKPDIIVTYGAGVRGSGVKGCGCRALSGTSVASPV 420
QY 421 AGAVTLVSTVQKRELNVNPSAKMALLASARRLPGVNMFEQHGKLDLLRAYOILNSYKP 480
DB 421 AGAVTLVSTVQKRELNVNPSAKMALLASARRLPGVNMFEQHGKLDLLRAYOILNSYKP 480
QY 481 QASLSFSYDILTECPYMWPCSOPIYYGMPYIIVNVLNMGVGTGRVYDKFENRYPYLPQ 540

DB 481 QASLSFSYDILTECPYMWPCSOPIYYGMPYIIVNVLNMGVGTGRVYDKFENRYPYLPQ 540
QY 541 NGDNTEVAFSSYSSVLPWWSGYLAISVTKKAAWEGIAOQHVMITVAPAEATESKNAGAE 600
DB 541 NGDNTEVAFSSYSSVLPWWSGYLAISVTKKAAWEGIAOQHVMITVAPAEATESKNAGAE 600
QY 601 QTSYKLPKIKYIIPTPPRSKRVLMDOYHNLRYPGYPFRDNLNRMKNDPLDWNQGDVHTN 660
DB 601 HTSTYKLPKIKYIIPTPPRSKRVLMDOYHNLRYPGYPFRDNLNRMKNDPLDWNQGDVHTN 660
QY 661 FRDMYQHLRSMGYVEVGLGAPFTCFDASQYGTLLMVDSEEEYFPEETIAKLRRDVGDLGL 720
DB 661 FRDMYQHLRSMGYVEVGLGAPFTCFDASQYGTLLMVDSEEEYFPEETIAKLRRDVGDLGL 720
QY 721 VIFSDWNTYSVARKVKFYDENTROMMPDTGGANIPALNELLSSVWNGFSDGLYGEFTL 780
DB 721 VIFSDWNTYSVARKVKFYDENTROMMPDTGGANIPALNELLSSVWNGFSDGLYGEFTL 780
QY 781 ANHDMYASGCSIAKPPEDGVVITOTFDQGLVLEKOTAVVENVPILGYOIPAEGGGR 840
DB 781 ANHDMYASGCSIAKPPEDGVVITOTFDQGLVLEKOTAVVENVPILGYOIPAEGGGR 840
QY 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900
DB 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900
QY 901 EGNHLHRYSKVLEAHLGDKPRPLPACPLSWAKPQPLNETAFSNLWKHOKLLSIDDKV 960
DB 901 EGNHLHRYSKVLEAHLGDKPRPLPACPLSWAKPQPLNETAFSNLWKHOKLLSIDDKV 960
QY 961 VLPNFRNRPOVRPLSPGSGAWDIPGGIMPGRYNQEVGOTIPVFAFLGAMVYVAFVYQ 1020
DB 961 VLPNFRNRPOVRPLSPGSGAWDIPGGIMPGRYNQEVGOTIPVFAFLGAMVYVAFVYQ 1020
QY 1021 INKAKSRPKRKRPRKPRQMLQOQVHPKPTPSV 1052
DB 1021 ISKAKSRPKRKRPRKPRQMLQOQVHPKPTPSV 1052

RESULT 11
ABB90255
ID ABB90255 standard; Protein; 666 AA.
XX ABB90255;
XX 24-MAY-2002 (first entry)
XX Human polypeptide SEQ ID NO 2631.
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX WO200190304-A2.
XX 29-NOV-2001.
XX 18-MAY-2001; 2001WO-US16450.
XX 19-MAY-2000; 2000US-205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Birse CE, Rosen CA;
XX WPI; 2002-122018/16.
XX N-PSDB; ABL90664.
XX

CC encoded proteins (see AAB20012-15) involved in the SREBP pathway.
CC C. elegans and D. melanogaster animals genetically modified to
CC express or mis-express these proteins are claimed. These
CC genetically modified animal models have identifiable phenotypes
CC that make them useful in assays for studying lipid metabolism,
CC other genes implicated in lipid metabolism and compounds capable
CC of modulating lipid biosynthetic pathways. Model organisms or
CC cultured cells can be used to identify new drug targets,
CC therapeutic agents, diagnostics and prognostics of disorders
CC associated with lipid metabolism, and also to identify pesticide
CC targets directed to components of the SREBP pathway.
XX
SQ

Sequence 992 AA;

Query Match 42.1%; Score 2366.5; DB 22; Length 992;
Best Local Similarity 46.6%; Pred. No. 1.4e-166;
Matches 466; Conservative 151; Mismatches 232; Indels 151; Gaps 16;
Qy 47 FSSVVEYEVIVAFNGYFTAKARNSFTSSALKSSVDNWRILIPNPNSSDYPSDFEVIQI 106
Db 19 FKTAVPNEFIVHFHFKYFAPVRESYIAAKLLGSNTWNRIVPRNLNLAQYPSDFDIIRV 78
Qy 107 KEQKAG----LTLTDHPNKRVTPOKVFRLSKY-RESPTVPCNTRWQKQSSRP 161
Db 79 CDGESSESFEIIRLQTHPSKAVVQPSVRRLINDAYSMLT----- 121
Qy 162 LRASLSLGSFGFWHATGRHSRRLLRAIP-----RQVAGTLQADVLQWQGYTCANRVAV 216
Db 122 -----YTH---RHQGVLRNPNNDHRQLC SVLHANILWKLGITGKGVKAI 167
Qy 217 FDTGLSEKHPHFNKERTNWTNERTLDDGLGHTGFVAGVIASMRCCQGFAPDAELHIFR 276
Db 168 FDTGLTKNHPHFNKERTNWTNEKSLDDRVSHGTFVAGVIASSRCLGAFAPDAELYFK 227
Qy 277 VETNQSVYTSWFLDAFNAYLKKIDVLNLSIGGDFDMDFPVKVMELTANNLWVSAI 336
Db 228 VETNSQSVYTSWFLDAFNAYLKKIDVLNLSIGGDFDMDFPVKVMELTANNLWVSAI 287
Qy 337 GNDGLYGLTLNPNADQMDVIGVGIGIDFEDNIAFSSRGMTTWELPGYGRMKPDVTYGA 396
Db 288 GNDGLYGLTLNPNADQMDVIGVGIGIDFEDNIAFSSRGMTTWELPGYGRMKPDVTYGA 347
Qy 397 GVRSGVKGCRALSGTSVSPVAVAGATVLLYS-TVQKRELNVNASKMOLIASARLLPG 455
Db 348 QVEGSDVYKCRRLSGTSVSPVAVAGAAALLISGAFQIDYINPASKLVLEGAKLPH 407
Qy 456 VNMFEQGHGKLDLLRAYQILNSYKPOASLSPSYDLTCEPYMWPYCSQPIYTGMPVTYN 515
Db 408 YNMFQAGKLLNLSMQLLSYKPKITLIPALYDFTQ-NYMMFYSQPIYTGSSVAIAN 466
Qy 516 VTILNGMVGTRIVDKPQWQPYLPQNGDNIEVAFYSVLWPNWSGYLAISISVTKKAASW 575
Db 467 VTILNGISVTSHTVIGPKWIPDFENQGLQVSAQVSPVWPMWGMVFIARKEGENF 526
Qy 576 EGIAQGHMVTIVASPAETESKNGAQTSTVKLPKIKVITPPPRSKRVLDQYHNLRYPP 635
Db 527 EGVCCKGSITLVLESFQKNTNET---HVTEDVPLIKVITPPPRSKRVLDQYHNLRYPP 583
Qy 636 GVFPDRNLKMDPLDNGDHIHTNFRDMYQHLRSMGVFEVLGAPTCFDSAQYGTLLM 695
Db 584 RYIPRDLKVKLPLDWRADHIHTNFRDMYHLRVNGYIDVLEPFTCFNADSGALLI 643
Qy 696 VDSSEEFPEETAKLRDY-DNGLSLVIFSDWNTSVMRKVKFYDENTQWMPDPTGGAN 754
Db 644 VDPERGGEDEINLAQENYVRGLNVVFGDMWNTVMKKIKFDFENTQWMPDPTGGAN 703
Qy 755 IPALNELLSVWNGFSDGLYEGETFLANHDMMYYAGSCSIKAFPD--GVVITQTFKDOGL 812
Db 704 IPALNLLKRFPGAFGDFVGEHFKLDHSMYYASGATVVKFPMNPDGIIVCTKLNDOGL 763
Qy 813 EYLKQET---AVVENVPILGLYQIPA----- 835
Db 764 SIINKTSPSKAKLDVPIFGMFQTKANSIQSNEEIVNVAESNLAEIPTDYSTFKNRVLL 823

Qy 836 -----EGGRIVLYGDSNCLDSDSHROKDCFMLDALLQYTSYGVTPP 877
Db 824 LTKORSISPAKSNHETKREGRIAYGDSNCLDSTHLEKACWLLITFLDFAIN----- 878
Qy 878 SLUSHGNQRPPSGAGSVTPERMENHLYRYSKYVLEAHLGDPKPRPLPA----- 926
Db 879 --SH-----KSSLQNI.NRTEPHKLERAPLPLRISOSITKRSQD 917
Qy 927 --CPLRSWAKPQPLNAPTAPSNLWKHQKLLSIDLDKVVLPN 964
Db 918 NNCEQFWLAPTQNNNA-----EERKSIIDVTILEN 949

RESULT 13

ABB63047

ID ABB63047 standard; Protein; 952 AA.

XX ABB63047;

XX AC

XX XX

XX DT 26-MAR-2002 (first entry)

XX XX

XX DE Drosophila melanogaster polypeptide SEQ ID NO 15933.

XX XX

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX XX

XX OS Drosophila melanogaster.

XX XX

XX PN WO200171042-A2.

XX XX

XX PD 27-SEP-2001.

XX XX

XX PF 23-MAR-2001; 2001WO-US09231.

XX XX

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX XX

XX PA (PEKE) PE CORP NY.

XX XX

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI: 2001-656860/75.

XX DR N-PSDB; ABL07150.

XX XX

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX PT interactions -

XX XX

XX PS Disclosure; SEQ ID NO 15933; 21pp + Sequence Listing; English.

XX XX

XX CC The invention relates to an isolated nucleic acid detection reagent

XX CC capable of detecting 1000 or more genes from Drosophila. The invention is

XX CC useful in developmental biology and in elucidating cell signalling and

XX CC cell-cell interactions in higher eukaryotes for the development of

XX CC insecticides, therapeutics and pharmaceutical drugs. The invention

XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX CC sequences (ABL01840-ABL16175) and the encoded proteins

XX CC (ABBS7737-ABBS72072).

XX CC The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX XX

XX SQ Sequence 952 AA;

Query Match 39.7%; Score 2231; DB 22; Length 952;

Best Local Similarity 44.5%; Pred. NO. 1.9e-156;

Matches 445; Conservative 147; Mismatches 230; Indels 178; Gaps 17;

Qy 47 FSSVVEYEVIVAFNGYFTAKARNSFTSSALKSSVDNWRILIPNPNSSDYPSDFEVIQI 106

Db 19 FKTAVPNEFIVHFHFKYFAPVRESYIAAKLLGSNTWNRIVPRNLNLAQYPSDFDIIRV 78

FT Misc-difference 265 /note= "Ser or Leu"
XX US5719021-A.
XX 17-FEB-1998.
XX 31-JUL-1992; 92US-0923260.
XX 02-MAY-1989; 89US-0346552.
XX (UYNE-) UNIV NEW JERSEY.
XX Inouye M;
XX WPI; 1998-158792/14.
XX Activation of polypeptides - by interaction with activating peptide,
XX resulting in refolding of the polypeptides to give active form
XX Disclosure; Columns 39-42; 29pp; English.
XX The present sequence represents pro-subtilisin of Bacillus
XX subtilis, and was used to demonstrate the method of the invention.
XX The propeptide is essential for the production of active, correctly
XX folded subtilisin. The M18T substitution resulted in a suppression
XX of this activity. An in vitro method to restore or increase the
XX natural biological activity of a target polypeptide (inactive or with
XX decreased activity due to improper folding), which is normally expressed
XX containing a prosequence forms the basis of the invention. An exogenous
XX activating peptide used to promote refolding of the target polypeptide
XX to give its active form. The activating peptide comprises the prosequence
XX of the target or other proteins with a similar sequence and function to
XX the target polypeptide. The method is used to produce biologically,
XX correctly folded proteins from their inactive, incorrectly folded forms.
XX Suitable target polypeptides include members of the serine protease or
XX subtilisin families.
XX Sequence 352 AA;
Query Match 6.0%; Score 338.5; DB 19; Length 352;
Best Local Similarity 26.8%; Pred. No. 1e-16;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSTVVEYIVAFNGYFTAKARNSFISALKSEVDNWRIRPNNSDDYPSDFEVIQIK 107
DB 4 SST--EKKIYVGF-----KQTSAMSSAKKKDVI----- 30
QY 108 EKQAGLITLE-DHPNKKRVTPQKRVFSLKVAESDPTVPCNETRWQKWSRPLRRAS 166
DB 31 -SEKGGKVKQKQFYVNAATAATLDEKAVKELK---KDPVA----- 66
QY 167 LSLGSGFWHATGRHSRRLLRAIPROVAOTLQADVLQWGYTGANRVAVDFDGLSEKHP 226
DB 67 -----YVEDHIAHEVAGSVPGISQ-IPALHSGQYTGSNRVAVIDSGIDSSHP 117
QY 227 HFKNKERTWNTERT--LDGLGHGTFVAGVTASWRE---CQGFADPAELHIFRVFTNN 281
DB 118 DL-NVRGASGFVSETNPNYQDSSGSHGTHVAGTIALNNSIGVLGVSASLAVKVLDS 176
QY 282 QVSYTSWFLDAFNALIKKIDVNLISIGGPFMD--HPFDKRWELTANNVIMVSAIGND 339
DB 177 GSGQYIINGIEWAISSNNMDVNNISLGGPTGTSTALKTVVDKA---VSSGIYVAAAGNE 233
QY 340 GP--LYGTLLNPADQMDVIGGIDFEDNIAREFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
DB 234 GSSGSTGVTPYPAKYPSTIAGVAVNSNQRAFSSAGS-----EL-----DVNAPGV 281
QY 398 VRGSGVGGCRALSGTFSVASPWAGAVTLLVSTVQKRELNVNPSMKOALIASARRLPGVN 457
DB 282 IQSTLPGGTVGAVNGTSMATPHVAGAAALILS---KHPTWTNAQVRDLRESTATVIG--N 336
QY 458 MFEQGHCKLDDLRLAYQ 473

DB 337 SFYIGKGLINVQAAQ 352
RESULT 19
AAW46595
ID AAW46595 standard; protein; 352 AA.
XX AAW46595;
XX 22-MAY-1998 (first entry)
XX Amino acid sequence of pro-subtilisin E from Bacillus subtilis.
XX Pro-subtilisin; propeptide; subtilisin; subtilisin E; active;
XX folding; increase; activity; improper folding; prosequence;
XX activating peptide; serine protease.
XX Bacillus subtilis.
XX Key Location/Qualifiers
XX Region 1..77 /note= "propeptide"
XX US5719021-A.
XX 17-FEB-1998.
XX 31-JUL-1992; 92US-0923260.
XX 02-MAY-1989; 89US-0346552.
XX (UYNE-) UNIV NEW JERSEY.
XX Inouye M;
XX WPI; 1998-158792/14.
XX Activation of polypeptides - by interaction with activating peptide,
XX resulting in refolding of the polypeptides to give active form
XX Disclosure; Columns 25-28; 29pp; English.
XX The present sequence represents pro-subtilisin of Bacillus subtilis
XX strain 168, and was used to demonstrate the method of the invention. To
XX investigate the role of the propeptide of subtilisin, an expression
XX system for the active subtilisin E was established in Escherichia coli
XX by replacing the pro-sequence of pre-pro-subtilisin with the E. coli
XX OmpA signal peptide. When the amino terminal 14 residues of the
XX propeptides are not removed. The propeptide is essential for the
XX production of active, correctly folded subtilisin. An in vitro method
XX to restore or increase the natural biological activity of a target
XX polypeptide (inactive or with decreased activity due to improper
XX folding), which is normally expressed containing a prosequence forms the
XX basis of the invention. An exogenous activating peptide used to promote
XX refolding of the target polypeptide to give its active form. The
XX activating peptide comprises the prosequence of the target or other
XX proteins with a similar sequence and function to the target polypeptide.
XX The method is used to produce biologically, correctly folded proteins
XX from their inactive, incorrectly folded forms. Suitable target
XX polypeptides include members of the serine protease or subtilisin
XX families.
XX Sequence 352 AA;
Query Match 6.0%; Score 337.5; DB 19; Length 352;
Best Local Similarity 26.8%; Pred. No. 1.2e-16;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSTVVEYIVAFNGYFTAKARNSFISALKSEVDNWRIRPNNSDDYPSDFEVIQIK 107
DB 4 SST--EKKIYVGF-----KQTSAMSSAKKKDVI----- 30

FH Key Location/Qualifiers
 FT Modified-site 1
 XX EP357157-A.
 XX PD 07-MAR-1990.
 XX PF 22-JUN-1984; 84EP-0202584.
 XX PR 24-JUN-1983; 83US-0507419.
 XX PR 29-MAY-1984; 84US-0614612.
 XX PR 29-MAY-1984; 84US-0614615.
 XX PR 29-MAY-1984; 84US-0614616.
 XX PR 29-MAY-1984; 84US-0614617.
 XX PR 29-MAY-1984; 84US-0614491.
 XX PA (GETH) GENENTECH INC.
 XX PI Bott RR, Estell DA, Ferrari E, Henner DJ, Wells JA;
 XX DR WPI: 1990-069509/10.
 XX DR N-PSDB; AAQ03536.
 XX
 PT Mutant prokaryotic carbonyl hydrolase enzymes -
 PT obtd. by site-directed oligo-nucleotide mutagenesis, used in
 PT food processing and cleaning industries.
 PS Claim 16; Fig 7; 39pp; English.
 XX
 CC Probe derived from subtilisin gene was used to isolate carbonyl
 CC hydrolase gene, mutant versions of which exhibit different oxidative
 CC stability and/or pH activity.
 XX
 SQ Sequence 381 AA;
 Query Match 6.0%; Score 337.5; DB 11; Length 381;
 Best Local Similarity 26.8%; Pred. No. 1.4e-16;
 Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
 QY 48 SSTVVEYEVVAENGFTAKARNSTFISSALKSEVDNWRRIIPRNPSSDYPDFEVIQIK 107
 DB 33 SST--EKYIVGF-----KOTMSAMSSAKKDKVI----- 59
 QY 108 EKQKAGLLTLE-DHPNKKRVTPQKVFSLKYAESDPTVPCNETRWSQKWSRPLRRAS 166
 DB 60 -SEKGGKVKQKFKYVNAATAATLDEKAVKELK---KDPVA----- 95
 QY 167 LSLSGGFHATGHRSSRLRLRAIPROVAOTLQADVLQMGYTGANVRVAFDTGLSEKHP 226
 DB 96 -----YVEEDHIAHEYAQSVPGISQ-IPALHSGQGTGNSVAVVDSGIDSSHP 146
 QY 227 HFKNVKERTNWTNERT--LDDGLGHGTFVAGVIASMR-----COGFAPDAELHIFRVFTNN 281
 DB 147 DL-NVRGASFPVSETPNPYQDSSGHGTHVAGTIAALNNSIGVLGVSFASLYAVKYLDT 205
 QY 282 QVSYTSFLDAFNALKKIDVNLISGGPDFMD--HPFVDKVVWELTANNVIMVSAIGND 339
 DB 206 GSGQYSWINGIEWAINNNMDVINNSLGGPTGTALKTVVDKA---VSSGIVVAAAGNE 262
 QY 340 GP--LYGTLANPADQMDVIGGIDFEDNIARESSRGMTTWELPGCYGRMKPDIVTYGAG 397
 DB 263 GSSGSTVGYPKAPYPTIAGVAVNSNQASFSAGS---EL-----DVNAPGVS 310
 QY 398 VRGSGVGGCRALSGTSPASVPWAGAVTLLVSTVQKRELVPNSMKOALITASARLPGVN 457
 DB 311 IQSTLPGGTGYGNGTSMATPHVAGAAALILS---KHFTWNAQVRDLSESTATYLG--N 365
 QY 458 MPEQGHCKLQDLRAYO 473
 DB 366 SFYTGKGLINVOAAQ 381

RESULT 22
 AAR24131
 ID AAR24131 standard; Protein; 381 AA.
 XX
 AC AAR24131;
 XX
 DT 17-NOV-1992 (first entry)
 XX
 DE Bacterial serine protease mutant.
 XX
 KW Detergent; liquid; lipase; degradation; stable.
 XX
 OS Humicola lanuginosa.
 XX
 FH Key Location/Qualifiers
 FT Cleavage-site 106..107
 FT Protein /note="cleaves signal peptide"
 FT 107..381
 FT /note="mature protease"
 XX
 XX EP486073-A.
 XX PD 20-MAY-1992.
 XX
 XX PF 25-JAN-1991; 91EP-0200149.
 XX PR 14-NOV-1990; 90EP-0870212.
 XX PR 25-JAN-1991; 91EP-0200149.
 XX
 XX (PROC) PROCTER & GAMBLE CO.
 XX
 XX Johnston JP, Lenoir PMA, Thoen CAJK, Thoen CAJ;
 XX WPI: 1992-168685/21.
 XX
 PT Liquid detergent compsn. contg. lipase and protease - i.e.
 PT bacterial serine protease in which methionine near active site is
 PT replaced
 XX
 PS Disclosure; Page 9; 10pp; English.
 XX
 CC The sequence is that of a bacterial serine protease in which the
 CC methionine adjacent to the serine of the active site is replaced by
 CC another amino acid. The protease is used in a detergent
 CC composition which comprises conventional detergent ingredients plus
 CC an enzyme system comprising a lipase and the protease variant. The
 CC protease variant does not attack the lipase and thus prevents lipase
 CC degradation. Liquid detergent compsns. are obtd. which are stable
 CC during storage.
 XX
 SQ Sequence 381 AA;
 Query Match 6.0%; Score 337.5; DB 13; Length 381;
 Best Local Similarity 26.8%; Pred. No. 1.4e-16;
 Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
 QY 48 SSTVVEYEVVAENGFTAKARNSTFISSALKSEVDNWRRIIPRNPSSDYPDFEVIQIK 107
 DB 33 SST--EKYIVGF-----KOTMSAMSSAKKDKVI----- 59
 QY 108 EKQKAGLLTLE-DHPNKKRVTPQKVFSLKYAESDPTVPCNETRWSQKWSRPLRRAS 166
 DB 60 -SEKGGKVKQKFKYVNAATAATLDEKAVKELK---KDPVA----- 95
 QY 167 LSLSGGFHATGHRSSRLRLRAIPROVAOTLQADVLQMGYTGANVRVAFDTGLSEKHP 226
 DB 96 -----YVEEDHIAHEYAQSVPGISQ-IPALHSGQGTGNSVAVVDSGIDSSHP 146
 QY 227 HFKNVKERTNWTNERT--LDDGLGHGTFVAGVIASMR-----COGFAPDAELHIFRVFTNN 281
 DB 147 DL-NVRGASFPVSETPNPYQDSSGHGTHVAGTIAALNNSIGVLGVSFASLYAVKYLDT 205
 QY 282 QVSYTSFLDAFNALKKIDVNLISGGPDFMD--HPFVDKVVWELTANNVIMVSAIGND 339

Db 206 GSGQYSWIINGIEWALSNNMDYINMSLGGTGTALKTVVDKA---VSSGIIVVAAAAGNE 262
QY 340 GP--LYGTNNPADQMDVIGVGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
Db 263 GSGGSTSTGVYPAKYPSTTAVGAVNSNQRASFSSAGS---EL-----DVMAPGVS 310
QY 398 VRGSGYKGCRLSGTSVASPVVAGAVTLLVSTVQKRELVPASMKQALIASARRLPVGN 457
Db 311 IQSTLPGGYTGAYNGTSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--N 365
QY 458 MFEQGHGKLDLLRAYQ 473
Db 366 SFYKGLINVOQAAAQ 381

RESULT 23

AAR34463
ID AAR34463 standard; Protein; 381 AA.

AC AAR34463;

DT 19-AUG-1993 (first entry)

DE Bacillus subtilis alkaline elastase.

XX Alkaline protease; detergent; stability; Bacillus; bleach; stains;
KW alkylbenzene sulphonates; cheese manufacture; dehairing enzyme;
KW leather preparation.

XX Bacillus subtilis.

XX W09307276-A.

PN 15-APR-1993.

PD 07-OCT-1992; 92W0-0508341.

PF 08-OCT-1991; 91US-0772087.

PR (CHEM-) CHEMGEN CORP.

XX (VIST-) VISTA CHEMICAL CO.

PI Fodge DW, Hsiao H, Lalonde JJ;

XX WPI; 1993-134465/16.

XX Alkaline protease produced by bacillus stable in alkaline
PT conditions - used in detergents and bleaches to decompose
PT proteinaceous stains

PS Disclosure; Page 34; 58pp; English.

XX Bacteria producing an alkaline protease that is stable at a pH
CC greater than 10 in the presence of linear alkylbenzene sulphonate
CC were isolated by collecting soil from an alkaline environment which
CC had been exposed to detergent contamination or which had a pH greater
CC than 12. Strains were isolated which had alkaline protease activity
CC which was stable in the presence of detergent. The protein
CC sequence was compared to that of Bacillus species alkaline elastases,
CC from B. carlsberg, B. subtilis and B. amyloliquefaciens BPN to
CC determine if the novel protease was from a bacillus species. The
CC novel protease showed between 60-75 percent homology to the elastase
CC from the other strains, thus indicating that the protease is from a
CC novel Bacillus strain designated Bacillus 164A. The alkaline protease
CC is used in heavy duty detergents as the enzymes are useful to decompose
CC proteinaceous stains and perform at high pH which are beneficial to the
CC detergency of the surfactants in the liquids. The enzyme is stable for
CC long periods in the detergent liquids and when stored at high pH. The
CC enzyme is also useful in cheese mfr. and as a dehairing enzyme for
CC leather preps. See also AAR34258-61 and AAR34463-6.

XX Sequence 381 AA;

(GEMV) GENENCOR INC.

Query Match 6.0%; Score 337.5; DB 14; Length 381;
Best Local Similarity 26.8%; Pred. No. 1.4e-16;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;

QY 48 SSTVVEYEVYAFNGYFTAKARNSFISSALKSSEVDNWRRIIPRNPSSDYPSPDFEVIQIK 107
Db 33 SST--BKKYIVGF-----KQTSAMSSAKKQDVI----- 59
QY 108 EKQKAGLLTLE-DHPNLIKRVTPQKVFYRSIKYAESDPTVPCNETRWSQKWQSSRPLRRAS 166
Db 60 -SEKGGKQKQFYVNAATAATLDEKAVKELK---KQPSVA----- 95
QY 167 LSLGSGFWHATGRHSRRRLRAIPROVAQTLOADVLQMCGYTGANVRVAYFDTGLSEKHP 226
Db 96 -----YVEEDHIAHEYAQSVYPYGISQ-IPALHSGQGYTGSNVKVAVIDSGIDSSHP 146
QY 227 HFNKVKERTNWTNERT--LDDGLGCHGTFFVAGVTASMR-----COGFAPDAELHIFRVPTNN 281
Db 147 DL-NVRGGASFVFPSETNPYQDSSGSHGTHVAGTTAALNNSIGVLGVSPSASLYAVKVLDT 205
QY 282 QVSYTSMELDAFNAYAILKKIDVLNLSIGGPDFMD--HPFVDKYWELTANNVIMVSAIGND 339
Db 206 GSGQYSWIINGIEWALSNNMDYINMSLGGTGTALKTVVDKA---VSSGIIVVAAAAGNE 262
QY 340 GP--LYGTNNPADQMDVIGVGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
Db 263 GSGGSTSTGVYPAKYPSTTAVGAVNSNQRASFSSAGS---EL-----DVMAPGVS 310
QY 398 VRGSGYKGCRLSGTSVASPVVAGAVTLLVSTVQKRELVPASMKQALIASARRLPVGN 457
Db 311 IQSTLPGGYTGAYNGTSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--N 365
QY 458 MFEQGHGKLDLLRAYQ 473
Db 366 SFYKGLINVOQAAAQ 381

RESULT 24

AAR74224
ID AAR74224 standard; Protein; 381 AA.

XX AAR74224;

DT 10-JAN-1996 (first entry)

DE B. subtilis subtilisin.

XX Bacillus amyloliquefaciens subtilisin; carbonyl hydrolase; protease;
KW fusion protein; preproprotein; transport; cell membrane; B.subtilis;
KW autolysis; maturation.

OS Bacillus subtilis.

Key	Location/Qualifiers
Peptide	1..106
FT	/label= signal peptide
FT	107..381
FT	/label= mature peptide

XX U05411873-A.

PN 02-MAY-1995.

XX 29-MAY-1984; 84US-0614612.

XX 01-APR-1986; 86US-0846627.

PR 29-MAY-1984; 84US-0614612.

PR 27-FEB-1990; 90US-0488433.

PR 11-AUG-1992; 92US-0928697.

XX (GEMV) GENENCOR INC.

XX

PI Adams RM, Power SD, Powers DB, Wells JA, Yansura DG;
 XX WPI; 1995-178127/23.
 DR N-PSDB; AAQ90042.
 XX
 PT Recovery of recombinant subtilisin mutants from host cells - by
 PT treatment with active subtilisin to cleave mutant from its
 PT pro-sequence.
 XX
 PS Disclosure; Fig 2; 32pp; English.
 XX
 CC The amino acid sequence of the Bacillus subtilis subtilisin protein.
 CC The gene is used in a method to produce a carbonyl hydrolase
 CC (subtilisin) e.g. the B.amyloliquefaciens subtilisin (AAQ90041) or other
 CC heterologous protein (produced as a fusion protein) e.g. human growth
 CC hormone, such that the desired protein is translated as a preproprotein
 CC which can be transported across the cell membrane but is not released as
 CC an enzymatically functional protein until the application of an external
 CC protease or a protease encoded by the host cell e.g. B.subtilis
 CC subtilisin or neutral protease (AAQ90043). The preproprotein sequence
 CC is mutated so that it is incapable of autolytic maturation. The
 CC B.amyloliquefaciens sequence was mutated using the primers AAQ90044-5
 CC and 0*****, specifically at S221N, D32N, A48R or contained a deletion
 CC of 166 amino acids from the C-terminus of the protein.
 XX
 SQ Sequence 381 AA;
 Query Match 6.0%; Score 337.5; DB 16; Length 381;
 Best Local Similarity 26.8%; Pred. No. 1.4e-16;
 Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
 QY 48 SSTVVEYIYAFNGYFTAKARNSFISALKSSVDNWRIRPNRNPSSDYPDFEVIQIK 107
 DB 33 SST--EKYIVGF-----KOTMSAMSSAKKKDVI----- 59
 QY 108 EQKAGLITLE-DHPNKRVTQPKVRSKLYAESDPTVPCNETRWSQKWQSSRLRRAS 166
 DB 60 -SEGGKVKQKQFYVNAATLDEKAVKELK---KDPVA----- 95
 QY 167 LSLGSGFWHATGKSHSRLLRAIPQVAOTLQADVLQWGYTGANRVAVFDLGLSEKHP 226
 DB 96 -----YVEEDHIAHEVAQSVFPGISQ-IPALHSGQYTGNSNVKVAVIDSGDSSHP 146
 QY 227 HFNKVKERTWNTERT--LDLGLGHGTFVAGVTASMBE---COGFADAEHLIFRFTNN 281
 DB 147 DL-NVRGASVFSEINPYQDSSGHGTHVAGTIALNNSIGVLGVSPASLAVKVLDT 205
 QY 282 QVSTYSWELDAFNAILKIDVLNLSIGGDFMD--HPFDKMWELTANNVIMVSAIGND 339
 DB 206 GSGQSWILNGIEWALSNMDVINNSLGGTGTSTALKTVYDKA---VSSGIIVVAAAAGN 262
 QY 340 GP--LYGTLANPADQMDVIGVGGIDFEDNITAFSSRGMTTWELPGYGRMKPDIYVYAG 397
 DB 263 GSGGSTVGYPAKYPTIAGVYNSNQRASFSSAGS---EL-----DVMAPGV 310
 QY 398 VRGSGVGGCRALSGTSVSPVAGAVTLVSVQKRELVPNSMKAQITASARLPQVN 457
 DB 311 IGSTLPGGTGYGNTGSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATVILG--N 365
 QY 458 MFEQGHGKLLRAYQ 473
 DB 366 SFYTGKGLINVAQAQ 381
 RESULT 25
 ID AAY39229 standard; Protein; 381 AA.
 XX
 AC AAY39229;
 XX
 DT 23-NOV-1999 (first entry)
 XX
 DE Bacillus subtilis subtilisin.

XX Subtilisin; protease; secreted protein; degraded; heterologous protein;
 KW catalytic triad mutation; autolytic maturation; prosequence;
 KW fusion protein.
 XX
 OS Bacillus subtilis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..106
 FT Protein /note="Putative signal peptide"
 FT 107..381
 FT /label= Subtilisin
 PN US5939315-A.
 PD 17-AUG-1999.
 XX
 PF 01-MAY-1995; 95US-04322279.
 PR 01-APR-1986; 86US-0846627.
 PR 29-MAY-1984; 84US-0614612.
 PR 27-FEB-1990; 90US-0488433.
 PR 11-AUG-1992; 92US-0928697.
 PR 01-MAY-1995; 95US-04322279.
 XX
 PA (ADAM/) ADAMS R M.
 PA (POWE/) POWER S D.
 PA (POWE/) POWERS D B.
 PA (WELL/) WELLS J A.
 PI Adams RM, Power SD, Powers DB, Wells JA, Yansura DG;
 DR WPI; 1999-526118/44.
 DR N-PSDB; AA206712.
 XX
 PT DNA sequence comprising subtilisin prosequence fused to a
 PT heterologous sequence or a catalytically inactive subtilisin
 PT variant, useful for the production of heterologous proteins
 XX
 PS Disclosure; Fig 2; 30pp; English.
 CC This sequence is the subtilisin protein of Bacillus subtilis. Subtilisin
 CC is an alkaline protease which Bacillus secretes. The protein sequence is
 CC used in the formation of a fusion protein which has an N-terminal
 CC subtilisin prosequence linked to a C-terminal non-subtilisin polypeptide.
 CC The C-terminal sequence can alternatively be a subtilisin which has an
 CC amino acid substitution in one or more of the catalytic triad of this
 CC sequence. The mutations take the form of a substitution at position 221
 CC in the wild type mature protein, where Ser is replaced by Ala. A
 CC substitution at position 32 in the wild type mature protein may also be
 CC used, with the wild type Asp replaced by Asn. A substitution may
 CC alternatively be made at position 64. These mutant subtilisins can be
 CC used as the second part of the fusion protein. The mutated versions of
 CC subtilisin are incapable of autolytic maturation. The DNA fusion
 CC sequences are useful for the production of heterologous proteins. As the
 CC subtilisin mutants are incapable of autolytic maturation, they are
 CC therefore bound to the Bacillus cell membrane. The inability of the
 CC mutant protein to leave the cell membrane means that they are not able to
 CC degrade desired proteins when heterologous protein production is being
 CC carried out in Bacillus cells. Heterologous proteins are sequestered at
 CC cell membranes and can be easily isolated and released by the action of
 CC enzymatically active subtilisin.

Sequence 381 AA;
 Query Match 6.0%; Score 337.5; DB 20; Length 381;
 Best Local Similarity 26.8%; Pred. No. 1.4e-16;
 Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
 QY 48 SSTVVEYIYAFNGYFTAKARNSFISALKSSVDNWRIRPNRNPSSDYPDFEVIQIK 107
 DB 33 SST--EKYIVGF-----KOTMSAMSSAKKKDVI----- 59

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:27:53 ; Search time 100 seconds
(without alignments)
2167.616 Million cell updates/sec

Title: US-09-830-837-6
Perfect score: 5617
Sequence: 1 MKLVNIWLLVLLVLCGKKH.....PRVKRPLMQVHPKPTPSV 1052

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2915	51.9	552	4 Q8TAN4	Q8tan4 homo sapien
2	2231	39.7	952	5 Q9VP10	Q9vp10 drosophila
3	335	6.3	1239	16 Q9FB24	Q9fb24 streptomyce
4	346.5	6.2	1722	5 Q9ND22	Q9nd22 leishmania
5	337.5	6.0	379	2 Q9FDF4	Q9fdf4 bacillus li
6	336	6.0	1102	2 P95684	P95684 streptomyce
7	335.5	6.0	1245	16 Q9RL54	Q9rl54 streptomyce
8	327.5	5.8	627	16 Q9RUD0	Q9rud0 deinococcus
9	324.5	5.8	379	2 Q45300	Q45300 bacillus li
10	322.5	5.7	374	2 Q9F942	Q9f942 bacillus li
11	322	5.7	382	2 Q45522	Q45522 bacillus sp
12	321	5.7	379	2 Q45301	Q45301 bacillus li
13	320.5	5.7	275	2 Q93166	Q93166 bacillus su
14	318.5	5.7	374	2 Q9F943	Q9f943 bacillus li
15	318.5	5.7	376	2 Q44684	Q44684 bacillus am
16	317.5	5.7	379	2 Q53521	Q53521 bacillus li

17	316.5	5.6	374	2	Q9F941	Q9f941 bacillus li
18	314.5	5.6	312	2	Q93Q20	Q93q20 clostridium
19	313	5.6	310	2	Q9FDF3	Q9fdf3 bacillus li
20	313	5.6	382	2	Q87655	Q87655 bacillus su
21	310.5	5.5	1253	16	Q9FC06	Q9fc06 streptomyce
22	308.5	5.5	275	2	Q9R7J4	Q9r7j4 bacillus su
23	308	5.5	310	2	Q9FDF2	Q9fdf2 bacillus li
24	304.5	5.4	323	2	Q45621	Q45621 bacillus sp
25	304	5.4	310	2	Q9F7C2	Q9f7c2 bacillus li
26	302	5.4	654	17	Q8U0C9	Q8u0c9 pyrococcus
27	297	5.3	412	2	Q9AER6	Q9aer6 thermoanaer
28	297	5.3	412	16	Q8RC68	Q8rc68 thermoanaer
29	294	5.2	321	2	P74937	P74937 thermoactin
30	293.5	5.2	322	16	Q9KAV3	Q9kav3 bacillus ha
31	291.5	5.2	640	2	Q934J3	Q934j3 prevotella
32	287.5	5.1	383	2	Q9KWR4	Q9kwr4 bacillus pu
33	287	5.1	379	2	Q45467	Q45467 bacillus sp
34	286.5	5.1	891	1	Q93635	Q93635 thermococcu
35	285.5	5.1	374	2	Q45523	Q45523 bacillus sp
36	285.5	5.1	379	2	Q45299	Q45299 bacillus li
37	285	5.1	1398	1	Q9P9L1	Q9p9l1 pyrococcus
38	284.5	5.1	561	16	Q8RBJ2	Q8rbj2 thermoanaer
39	278.5	5.0	379	2	Q66153	Q66153 bacillus sp
40	278.5	5.0	692	2	Q9EXK0	Q9exk0 bacillus ps
41	276.5	4.9	384	2	Q56365	Q56365 thermoactin
42	270.5	4.8	328	16	Q92Y41	Q92y41 rhizobium m
43	263.5	4.7	682	2	Q48674	Q48674 lactococcus
44	259	4.6	757	16	Q9K6G6	Q9k6g6 bacillus ha
45	258.5	4.6	397	2	P97097	P97097 bacillus sp

ALIGNMENTS

RESULT 1

ID	Q8TAN4	PRELIMINARY;	PRT;	552 AA.
AC	Q8TAN4			
DT	01-JUN-2002	(TReMBLrel. 21, Created)		
DT	01-JUN-2002	(TReMBLrel. 21, Last sequence update)		
DT	01-JUN-2002	(TReMBLrel. 21, Last annotation update)		
DE	Similar to membrane-bound transcription factor protease, site 1			
OS	Homo sapiens (human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
..RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RA	Strausberg R.;			
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC026330; A026330.1;			
KW	Protease.			
SQ	SEQUENCE. 552 AA; 61331 MW; FE28225645FFD6A8 CRC64;			

Query Match	51.9%;	Score 2915;	DB 4;	Length 552;
Best Local Similarity	99.8%;	Pred. No. 9.6e-205;		
Matches 551;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	MKLVNIWLLVLLVLCGKKHGLDRLEKKSFEKAPCGCSHLTLKVFSSVWEYIVAF	60	
Db	1	MKLVNIWLLVLLVLCGKKHGLDRLEKKSFEKAPCGCSHLTLKVFSSVWEYIVAF	60	
Qy	61	NGYFTAKARNSFTSSALKSEVDNWRLLPRNNSSDYPSPDFEVIQIKKQKAGLLLEHDH	120	
Db	61	NGYFTAKARNSFTSSALKSEVDNWRLLPRNNSSDYPSPDFEVIQIKKQKAGLLLEHDH	120	
Qy	121	PNIKRYTPQKVFPSRLKYAESDTPVPCNETRWQSQWSSRPLRRASLSLGSFGFWHATGRH	180	
Db	121	PNIKRYTPQKVFPSRLKYAESDTPVPCNETRWQSQWSSRPLRRASLSLGSFGFWHATGRH	180	
Qy	181	SSRRLLRAIPROVAQTLOADVLWQMGYTGANVAVFDTGLSEKHPHFKNKERTWNTE	240	
Db	181	SSRRLLRAIPROVAQTLOADVLWQMGYTGANVAVFDTGLSEKHPHFKNKERTWNTE	240	

```
Db 181 SSRRLRAIPROVAQFLQADVLQMGVGTGANVRKAVFDTGLSEKHFKNVKERTNWTNE 240
Qy 241 RTLDGLGHGHTFVAGVIASMRCEQGFAPDAELHIFRVTNNQVSYTSWFLDAPNVAIILK 300
Db 241 RTLDGLGHGHTFVAGVIASMRCEQGFAPDAELHIFRVTNNQVSYTSWFLDAPNVAIILK 300
Qy 301 IDVLNLSIGGPDMPDHPFVDRKMWELTANNVIMVSAIGNDGPLYGLTLNADQMDVIGVGG 360
Db 301 IDVLNLSIGGPDMPDHPFVDRKMWELTANNVIMVSAIGNDGPLYGLTLNADQMDVIGVGG 360
Qy 361 IDPENTARSSRGMTTWELPGYGRMKPDIVTYGAGVGRGSGVGGCRALSGTGSVASPW 420
Db 361 IDPENTARSSRGMTTWELPGYGRMKPDIVTYGAGVGRGSGVGGCRALSGTGSVASPW 420
Qy 421 AGAVTLVSVOKRELVPASMKQALIASARLPGVNMFEGHGHKDLRLAYOILNSYKP 480
Db 421 AGAVTLVSVOKRELVPASMKQALIASARLPGVNMFEGHGHKDLRLAYOILNSYKP 480
Qy 481 QASLSFSDIDLTCPYMWPCSOPIYVGMPTVYVNTILNGMGTGRIVDKPDQWPLPQ 540
Db 481 QASLSFSDIDLTCPYMWPCSOPIYVGMPTVYVNTILNGMGTGRIVDKPDQWPLPQ 540
Qy 541 NGDTEVAFSFS 552
Db 541 NGDTEVAFSFS 552

RESULT 2
Q9VP10 PRELIMINARY; PRT: 952 AA.
AC Q9VP10;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG7169 protein.
GN CG7169.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams R.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.J.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durlin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler G., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jastoli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
```

```
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL; AE003595; AAF51752.2; .
DR HSSP; P00782; 1SDP.
DR FlyBase; FBgn0037105; CG7169.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 952 AA; 107683 MW; 0391F95C89BD1D42 CRC64;

Query Match 39.7%; Score 2231; DB 5; Length 952;
Best Local Similarity 44.5%; Pred. No. 2.8e-154;
Matches 445; Conservative 147; Mismatches 230; Indels 178; Gaps 17;

Qy 47 FSSVTVEYEVIVAFNGYFTAKARNSFISALKSEVDNWRRIIPRNPSSDYPSPFVIOI 106
Db 19 FTAIVPNEFIVHPSKYFAPVRESYIAAKLGSNTNRIIVRLNLAQWQPSDFDLRV 78
Qy 107 KEKQAG-----LITLEDHFNIRKVTPOKVPERSLY-AESDPTVPCNETRWSQKSRP 161
Db 79 CDGYESSEFIERLOTDFSVAVVQPSVRRLINDAYSNTL----- 121
Qy 162 LRRASLSGGFWHATGRHSRRLLRAIP-----RQAQTLQADVLWQMTGANVAV 216
Db 122 -----YIH---RHQGVLRNPNNDNRRLQCSVLHANILKLGITGKGVKAI 167
Qy 217 FDTGLSEKHPHFKNVKNRTNWTNERTLDDGLGHGHTFVAGVIASMRCEQGFAPDAELHIFR 276
Db 168 FDTGLTKNHPFNKVNKERTNWTNKSLLDDRYSHGTFVAG----- 206
Qy 277 VFTNQVSYTSWFLDAPNVAIILKIDVLNLSIGGPDMPDHPFVDRKMWELTANNVIMVSAI 336
Db 207 -----VSYTSWFLDAPNVAIIRKINILNLSIGGPDMPDHPFVDRKMWELTANNVIMVSAI 260
Qy 337 GNDGPLYGLTLNADQMDVIGVGGIDFEDNIAREFSSRGMTTWELPGYGRMKPDIVTYGA 396
Db 261 GNDGPLYGLTLNADQMDVIGVGGIDFEDNIAREFSSRGMTTWELPGYGRMKPDIVTYGA 320
Qy 397 GYRGSYKGGCRALSGTGSVASPVAGVATLWS-TVQKRELVPASMKQALIASARLP 455
Db 321 QYEGSDVRKGCRRLSGTSVSPVAGAAALLISGAFQKIDYINPASLQVLEGAELPH 380
Qy 456 VNMFEQGHGKDLRLAYOILNSYKQASISPSYIDLTCPYMWPCSOPIYVGMPTVYN 515
Db 381 YNMFQAGAKLNLKSNQLLSYKPKITLIPALDFTQ-NYMWPTSSQPLYYGSSVAIA 439
Qy 516 VTILNGMGTGRIVDKPDQWPLPQNDGNIENAVSYSVLWPMWSGYLAISVTKKAASW 575
Db 440 VTILNGISVTHSHVIGIPKPIPEFENQGFQVSAQVSPVWPMWGMVSVFAVKEGENF 499
Qy 576 EGIAGQHVMTVAPASPTESKNGAEGTSTVKLPKIKVLIITPPRSKRVLNDOYHNLRYPP 635
Db 500 EGVCXGSHVTVLESFKQNET---HVTEDFPITIKVTPKPRNRIILMDQHSKRPP 556
Qy 636 GYFPRDLNLRMKNDPLDMNGDHTHTFRDMYQHLRSMGYFVEVIGCAPTCFDASQYGLIM 695
Db 557 RYTPRDDLLKVLDPDLWRADHHTNFRDMYTHLRNVGYIDVIREPFTCFNADSYGALLI 616
Qy 696 VDSSEYEPFEEIAKLREDV-DNGLSLVIFSDWYNTSVNRKVKFYDENTQWMPMDTGGAN 754
Db 617 VDPERGFDEEINALQENYKRGVNVVFGDWYNTTWKKIKKFFDENTQWMPMDTGGAN 676
Qy 755 IPALNELLSVNMWGFSDGLYEGETLHNDMYTASGCSIAKFPED--GVVITQTFKQDGL 812
```


Db 677 IPALNDLLKPGFAGDFVGEHFKLGDHSMYASGATIVKFPMPGDIIVGTKLNDQGL 736
QY 813 EVLKQET---AVVENPILGLYQIPA----- 835
Db 737 SIINSKTPSKVAKLDVDFGFWQKANSIQSNEEIVVNAESNLAEAIPTDYTFKNRVL 796
QY 836 -----EGGRIVLYGDSNCLDSDSHRQKDCFWLLDALLQYTSYGVTTP 877
Db 797 LRTKQRSISPAKSNHETKNEGRIVAVYGDNSCLDSTHLEKACYWLLITFLDFAIN----- 851
QY 878 SLSHSGNRQRPSPGAGSVTERMEGNHLHRYKSVLEAHLGDPKPRPLPA----- 926
Db 852 --SH-----KSLQLNLNRTFHKLERAPLRLISQSIKRSQD 890
QY 927 --CPRLSWAKPQPLNETAPSNLWKHOKLLSIDLKVVLN 964
Db 891 NNCEQPKWLAPTQKNA-----ERKSSIIDVILEN 922

RESULT 3
Q9FBZ4 PRELIMINARY; PRT; 1239 AA.
AC Q9FBZ4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative secreted peptidase.
GN SCO7188 OR SC8A11.16C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Lark L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL391041; CAPC01588.1; -.
DR HSSP; Q99405; IMPT.
DR InterPro; IPR001317; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 1239 AA; 128505 MW; 8F5E9AC68EB1260A CRC64;
Query Match 6.3%; Score 355; DB 16; Length 1239;
Best Local Similarity 22.8%; Pred. No. 9.3e-17;
Matches 217; Conservative 116; Mismatches 363; Indels 254; Gaps 45;
QY 203 WQMGYTCANRVAVDFDGLSEKHPFKN-VKERTNWTNERTLDGLGHGTFVAGVI----- 257
Db 236 WAGGNTGGVEAVLDGTGDAGHPDLADRIAAROSFVDPDENTDDRDGHGHVASTIAGTG 295
QY 258 -ASMRECGQFAPDAELHIFRVFTNNQVSYTSWFLDAFNALIKK-IDVLNLSIGCPDFMD 315
Db 296 AASAKKEGVAPGARLSIGKVLDSNGRQISWTLAAMEWAVERHAKIVNLSGSGQSD 355
QY 316 --HPFVDKVMWELTA-NNVIMVSAIGNDGPVYGLTNNPADQMDVITGVGGIDFEDNIARFSS 372
Db 356 GSDPMSRAVDRLSAQTGALFVVAAGNGEA-GSIGAPGVATSAITVGAVDATDTLAPFSS 414
QY 373 RGMTTWELPGYGRMKPDIVTYGAGVR-----GSGVKGGRALSCTSVASPVVAGAVTL 426
Db 415 QG-----PRVDGALKPEITAPGVGILAAANSFAAGNGAYQSLSGTSMATPHVAGAAAL 468
QY 427 LVSTVQKRELNVNPSMKQALIASARRLPGVNMFQGHGKLDL---LRA--YQILNSYKPK 481
Db 469 LAAA---RPDLSGSALKDVLASSHRTPRYDAFOAGSGRVDVDAVRAGVYASATAPG 525
QY 482 ASLSPSYIDLTECPYMPYCSQPIYGGMPTVNVNVTILNGMGVTRIGRVDKPDQWQYLPQN 541
Db 526 SSPG-----VRLVYTYTNTGAATLLELSVAATHAP----- 557
QY 542 GDNLEAVF--SYSSVLMPWSGYLAISVTKKAASWEGIAGGHVMTIVASPAETESKGA 599
Db 558 ----EGVRLSASRVTVPAHGTDVTLTIDGSGSAGGRAYSGQILAT-----DADARVA 608
QY 600 EOTSIVKLPIKVKII-----PTPPSRKRVLWDQYHNLRYPPG-----YFPRDN- 642
Db 609 -HTAVSAGPVVRHKLTVHFKDADGNPVGVDLLKSGDSESLPVLVGDSGTAELYPEDTY 667
QY 643 --LRMKNDPLDWDGDIHTNFRDMYQHLRSMGYFEVVLGAP-----FTCFDASQYGT 692
Db 668 SALAFKTVP-GVHGPH-----SWGMALLGDPEVRLTEDTAVTFDASR--- 708
QY 693 LLMWDSSEYFPE--ETAKLRDNDGLSLVIFSDWNTSVMRKVFVDENTRWMPDT 750
Db 709 ---VERIETTVPQTEATYQRLDYQRSWG-----GTYRTGLETOA-YDS---LWAQPTT 757
QY 751 GGA-----NIPALNELLVNMNMGFSGLYEGETLANHDMYASGSGSIARFP 797
Db 758 HKYTHGDFLYNARWRKEQPALT--VSTRTDFTDVLQSG-----GVTPALP 800
QY 798 EDGVVITQTFKDOGLE-----VLKQETAVVENPILGLYQIPAGGGGRIVLY 844
Db 801 KGTRTLPLVFAGDAAAEYARLARGAKAVVVRDDVDAD-----GVQAAANVAAGATILL 855
QY 845 GDSNCLDSDSHRQKDCFWLLDALLQYTSYGVTTPPSLSHSGNRQRPSPGAGSVTERMEGNH 904
Db 856 VVNN--EDGR-----ALRCY-----GEPFGPPVALDVALLSTDEGEK 890
QY 905 LHRYSKYLEAHLGDPKPRPLPACPRL-----SWAKPQPLNETAPSNLWKHOKLLSIDL- 957
Db 891 LAAQAKYRGARV---TVTSRPVSPYVYDILLASWINEIFTRMTSRAD---SRSLARVDVAF 944
QY 958 -----DKVVLNFRSNRPQVRPLSPGSGAMWDIPGIMPRGYNOEVQGITP 1003
Db 945 DSQLPGSGGGEFRYDVPVGGSGTGGQPEPVS-GTRTDWVSTGDY---RWNNQ----- 994
QY 1004 VFAFLGAMVVLAFVVOIN-KAKSRPKRRPR-----VKRPOLMQOVHPP 1047
Db 995 --AYAGGV-----IYEIGAKTAYRPGSROSEWFGPVRPHLNDAYRSP 1036


```
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-OWU 1411T;
RA Goldstein G., Wagner E.K., Lee Y.E., Masliah E.S., Oh J.;
RT "Nucleotide sequence of kera gene encoding a keratinolytic protease of
RL Bacillus licheniformis OWU 1411T.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF282893; AG00492.1;
DR HSSP: P00780; ICSE.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE.ASP; 1.
DR PROSITE: PS00137; SUBTILASE.HIS; 1.
DR PROSITE: PS00138; SUBTILASE.SER; 1.
SQ SEQUENCE 379 AA; 38904 MW; DC65F6A93EBC69CC CRC64;

Query Match 6.0%; Score 337.5; DB 2; Length 379;
Best Local Similarity 25.7%; Pred. No. 2.5e-16;
Matches 126; Conservative 66; Mismatches 169; Indels 129; Gaps 17;

QY 1 MKLVNTWLLLVLLCGKKHLGRLEKKSFEKAPCPGCSHLTLKVFSSV-----V 52
Db 2 MRKKSFWLGLTALM-----LVFTMAFSDSASAAQPAKNV 36

QY 53 EYEYIVAF-NGYFTAKARNSFISALKSSVD-NWRIIPRNPSSDYPSPDFEVIQIKEQ 110
Db 37 EKDIYVGFSGVKVTASVKKDIRES--GGKVDQKQFRIINAAKALD----- 80

QY 111 KAGLLTLEDPNLIKRVTPORKVPSRLKVAESDPTVPCNETRWSQKWSSRPLRRASLSLG 170
Db 81 KEALEEVKNDPV-----AYVEED----- 99

QY 171 SGFWHATGRSSRLRLRAIPRQVAQLQADVLQMGYTGANVRVAVDFDGLSEKHPHEKN 230
Db 100 -----HVALAQAQTPYGT-PLIKADKVAQGYKGANVRVAVLDGTGIQASHPLNV 149

QY 231 VKERTNWTNRTLDDGLGHGTFVAGVIASMQEQ---GPAPDAELHIFRVFTNQVSYTS 287
Db 150 VGGASFVAGEAYTWDGNGHTHVAGTVAALDNTGVLGVAPNVSLVAVKLVNSGSGSYS 209

QY 288 WFLDAFNIAIKKIDVNLISIGGDFMD--HPFVDKRWELTANNVIMVSAIGNDGPI--Y 343
Db 210 GIVSGIEWATTNGMDVINSLGSPSGSTAMQAVDNAY---ARGVVVAAAGNSGSGNT 266

QY 344 GTLNNADQMDVIGVGIGDFEDNIARFSSRGWTTWELPGYGRMKPDIVTYGAGVRSV 403
Db 267 NTIGYPAKVDVIAVAGVDSNRSFSSVGA---EL-----EVMAPGAGVYSTYP 314

QY 404 KGCRLALSGTSVASPVVAGAVTLLVSTVOKRELNVNPSMKQALIASARRLPGVNMPEQGH 463
Db 315 TSYATLNGTSMASPHVAGAAALILS---KHPNLSASQVRNRLSSTATYLG---SSFYKG 369

QY 464 GKDLLLRAYQ 473
Db 370 GLINVEAAQ 379

RESULT 6
P95684 ID P95684 PRELIMINARY; PRT; 1102 AA.
AC P95684;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Subtilisin-like protease.
OS Streptomyces albogriseolus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1887;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-S-3253;
RX MEDLINE=97144528; PubMed=8990295;
RA Suzuki M., Taguchi S., Yamada S., Kojima S., Miura K., Momose H.;
RT "A novel member of the subtilisin-like protease family from
RL Streptomyces albogriseolus.";
RL J. Bacteriol. 179:430-438(1997).
DR EMBL: D83672; BAA12040.1;
DR HSSP: P00782; 2SBT.
DR InterPro: IPR002860; GH_BNR.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF02012; BNR; 2.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE.ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE.HIS; 1.
DR PROSITE: PS00138; SUBTILASE.SER; 1.
KW Protease.
SQ SEQUENCE 1102 AA; 114128 MW; F9E4AD2590FE59E CRC64;

Query Match 6.0%; Score 336; DB 2; Length 1102;
Best Local Similarity 22.7%; Pred. No. 1.9e-15;
Matches 164; Conservative 107; Mismatches 260; Indels 190; Gaps 33;

QY 203 WQMGYTGANVRVAVFDITGLSEKHPHF-KYKERTNMTNERTLDDGLGHGTFVAGVIA--- 258
Db 209 WEAGYDGKGVKIAVLDTGVDTATPDLKGQVTSKNTFTSAPTTGVDVVGCHTHVASIAAGTG 268

QY 259 --SMRECQGPAPDAELHIFRVFTNNQVSYTSWFLDAPNAILKIDVNLISIGGPDPMOH 316
Db 269 AOSKGYTKGVAPGAKLNGKVLDDAGFDGDDSGILAGMEWAAAGADIVNNSLGG---MDT 325

QY 317 PFVD---KVELTA-NNVIMVSAIGNDGLYGTFLNPNADQMDVIGVGIDFEDNIARFS 371
Db 326 PETDPLEAAVDKLSAEKGLIFALTAAGNEP--OSIGSPGSADSALTGVAVDDDKKLADFS 383

QY 372 SRGMTWELPGYGRMKPDIVTYGAGVRSVGK-----GCRALSGTSVASPV 419
Db 384 STGPRL-----GDGAVKPDLTAPGVDTAASAKGNDIAKEVGEKPCAGYMTISGTSMTAPH 438

QY 420 VAGAVILLVSTVOKRELNVNPSMKQALIASARRLPGVNMPEQGHGKLDILRAYQILNSYK 479
Db 439 VAGAAALL---KQHPPEWYAEKLGALTASTKD--GKYPPEQSGRVRQVDKAITQTVAIE 494

QY 480 PQASLSPSYIDLTECPYMWPC-----SOPYYGGMPTVNVNITLNGMGTGRIVDKPDW 534
Db 495 P-----VSLSEGVQOHPHADDKPVTKKLTNRNLGT-EDVILKLTSTATG----- 537

QY 535 QPYLPQNGDNIEVAF---SYSSVLMPSGYLAISISV-TKKAASWEGIAQGHVMTVASP 590
Db 538 -----PKGKAAPAGFTTGLASTLTVPANGTASVDVTDATRLGGAVDGTYSAVVATGAG 592

QY 591 A-----ETSKNGAETSVKLPKVIKVIPTPPRS-----KRVLMD 626
Db 593 SVRTAAAREVESYN-----VTLKVLDRSGKATANYMAYLSGLTGLGKDRSYA 641

QY 627 QYH-----NLRYPPGYFPRD-NLRMKNDPLDWDGDIHTNFRDMYQHLRSMGYFVEYVGA 680
Db 642 PYEADGAVSVRVPKGGYVLDASVILGADPTWRG-----ADWLAQ 681

QY 681 PFTCFDAQSYGTLMLVDSEEE-----YFPEEIAKLK-----RDVNGLSLVIFSDWYNTS 730
Db 682 P--KLDVTR-NTTVTVDAKRAKPVKVTVPCKAQAQFASADYTIETNDSAVSYG----- 732

QY 731 VMKRVFYDENTQWMPDPTGGANIPAL-----NELLSVNMWGFSDGLEGEFTLANHD 784
Db 733 -----WPLENYSGRSAHLGPGITNGTLSQQWNTHFSNGA-KAQYTA----- 773

QY 785 MYVAGSGCIAKFPEDGVVITQTFKDOGLEVLKQSTAVVENVPILGVIQIPAEGGGRIVLY 844
Db 774 ---ISGKVKRL---ATGYTRAFK-----AKEFATVQ-----VGMGAASGKKGAVTAF 816

QY 845 G 845
```

Db 817 G 817

RESULT 7

Q9RL54 PRELIMINARY; PRT; 1245 AA.

AC Q9RL54; 159 SRPLRASLSIGSFWH-----ATGRHSR-----RLLRAPROVAQTQADYLM 203

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Probable secreted peptidase.

GN SC00432 OR SCF51A.10

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces.

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1902;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Murphy L., Harris D.;

RA STRAIN=A3(2);

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Thomson N.R., Parkhill J., Barrell B.G., Denapaita D., Eichner A., Cullum J.,

RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RN SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,

RA Kinashi H., Hopwood D.A.;

RA MEDLINE=97000351; PubMed=8843436;

EX "A set of ordered cosmids and a detailed genetic and physical map for

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

RN [4]

RN SEQUENCE FROM N.A.

RC STRAIN=A3(2); / M145;

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,

RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete Streptomyces

RT coelicolor A3(2).";

RL Nature 417:141-147(2002).

DR EMBL; AL121596; CAB56662.1; -

DR HSP; P00782; 1SDE

DR InterPro; IPR003137; PA

DR InterPro; P000209; Peptidase_S8.

DR Pfam; P02225; PA; 1

DR Pfam; P00082; Peptidase_S8; 1.

DR PRINTS; P00723; SUBTILISIN

DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.

DR PROSITE; PS00136; SUBTILASE_SER; 1.

SQ SEQUENCE 1245 AA; 130895 MW; 74EE92DB9C4DE60 CRC64;

Query Match 6.0%; Score 335.5; DB 16; Length 1245;

Best Local Similarity 21.5%; Pred. No. 2.5e-15;

Matches 218; Conservative 116; Mismatches 350; Indels 331; Gaps 42;

QY 159 SRPLRASLSIGSFWH-----ATGRHSR-----RLLRAPROVAQTQADYLM 203

Db 169 ARAVVDDEALGR-FWKRLVPAGGQAKAAATPRVWLDGRVGSASLDRSPAQIGAPDV-W 226

QY 204 QMGYTGANRVAVFDITGLSEKHPHF-KVKERTNMTNERTLDDGLGHGTFAVGI----- 257

Db 227 SAGYRGDGVKVAVLDTGADQSHPLAGRAAAKDFSGSGSGTNDVFHGCHTHVASIVGGSGA 286

QY 258 ASKRECQGFAPDAELHIFRVFTTNQVSYTSWFLDAFNAILKIDVLNLSIGGDPFMD-- 315

Db 287 ASGSRQGVAPAKRLVGVKVGDDGSESVIAGMEWAADQAGADVNNMSLGSAGTGT 346

QY 316 HFFVDKWEFT-ANNVIMVSAIGND---GPLYGLNNPQDMQDVGIVGGIDEDFNIAFS 371

Db 347 DPMQALNLSRRRTGTLFVVAAGNEGEGP--RTVSGPGAADAALTVGAVDRDLSAPFS 404

QY 372 SRGMTTWELPGGTGRMKPDIVTYGAGVSGVKGCC-----RALSGTSVASPVVAG 422

Db 405 SRGPRL-----GDDAVKPDVTAGVGIVAAARAAGSAMGDPVDEHYHTAASGTSMATPHVAG 459

QY 423 AVTLIVSTVQKRELVPASNKQALIASARRLPGVNMFEQGHGKLDLLRA----- 471

Db 460 AALL---AORHPDWTGAOLKDALISTAVTDGQKVTEQGGRIDVRAAGLAVTATGTL 516

QY 472 -----YQILNSYKPAQL----- 484

Db 517 VMGPFTSRDTEPVTSRVVNTSGSDEDVTLSLAVALATEGGKAPAGSARLSGDSVRVPAG 576

QY 485 SPYSIDLTECPYMPYCPSPIYGGMPVWVNTLNGM-----CVTGRI 528

Db 577 SSAEVLTVDP---ARAGQKFG---YVTATTADGSVAHAHTTSLVVGHPHTRITVRT 629

QY 529 VDKPDWQPLPQNGDNIEVAFSYSSVLAPWSYLAIISVTYKKAASWEGIAGHVMITVA 588

Db 630 IDK-----DGEQV-----ADLPTINGAEFGVGYSDPAAVAEVEEGYQLNTSWT-- 674

QY 589 SFAETESKNGAEQTSVKLPKV-----KIPTPPRSKRVKLVNDYHNLYPPGY 637

Db 675 -----TSAEDGELRHVLPVKVTKMTVTLDAKTVFVEITPR----- 715

QY 638 FPDNLNMKNDDPDMNGDHIHNFND-----MYQHLRSMGFVEV----- 677

Db 716 -----PABQRGILSYQTYREIGRSLTGTMYFDIAKRLYISITAKVTEGTFEF 764

QY 678 -----LGAPF-----TCFDASQY-----GTLIMVDS----- 698

Db 765 ASRWOLVAPLEAKVSRVSGTGLGAYIMPASLPPERGAITAVDAGVAADAFSRRGRV 824

QY 699 -----EEYEPPEIARLRDNGLSLVIFSDYNTSVMRKVKFYDENTROMMPTDGGAN- 754

Db 825 AVVDNPTGAGERALVERAAAGVSAILVHENDTGWTR-----WRP-TGDRNA 871

QY 755 IPALN-----ELLSVNMNFGSDGLYEGEFTLANHDMYYASGCSIAKFP-----E 798

Db 872 LPTLRIGKGTGAELLARIAGSTSVRFSG--TARSPYLVDVMQVSEQVPRKVVHSVSAR 929

QY 799 DGWVITQFKDQGLEVLKQETAV-----VENVPILGLYOIPAEGGGR--IVLYGDSNCLDD 852

Db 930 NSAVVRTTYADNGVTGWASEQRFGRWPYQNTAWLQYTRFVPLGFERTVSVSAGDTAMQHR 989

QY 853 SHROKDCFWLLDALLOYTSYGVTPPSLS-----HSGNQ-----RPSGAGS 894

Db 990 VHHT-----TTFDQVQPLIAGMRDPRTYAGVRPRETWOGAVVRPSIPAGT 1036

QY 895 VTPERMENGLH-RYSKYLEAHLGDPKRPPLPACPRLSWAKPOPLNETAPSILAK 948

Db 1037 TPTVTDGNNLRLVRAEFTDSQAG-----HNSPFGS-GDTHASAVLYR 1077

RESULT 8

Q9RUDDO PRELIMINARY; PRT; 627 AA.

AC Q9RUDDO; 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Serine protease, subtilase family.

GN DR1459.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;


```

DE Subtilisin precursor (EC 3.4.21.62) (Fragment).
GN SUBC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 10689;
RX MEDLINE=20561724; PubMed=11109488;
RA Evans K.L., Crowder J.S., Miller E.S.;
RT "Subtilisins of Bacillus spp. hydrolyze keratin and allow growth on
feathers.";
RL Can. J. Microbiol. 46:1004-1011(2000).
DR EMBL; AF205190; AAG31027.1; -.
DR HSSP; P00780; ICSE.
DR InterPro; IPR00209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Signal.
FT SIGNAL. 29 POTENTIAL.
FT CHAIN 30 >374 SUBTILISIN.
FT NON-TER 374
SQ SEQUENCE 374 AA; 38390 MW; B61432A1B42C526E CRC64;

Query Match 5.7%; Score 322.5; DB 2; Length 374;
Best Local Similarity 25.2%; Pred. No. 3.1e-15;
Matches 122; Conservative 66; Mismatches 168; Indels 129; Gaps 17;

QY 1 MKLVNILLVLLCGKHLGDLREKSKFAKPCGCSHLTKVFEFSIV-----V 52
   | : | | : | : |
   2 MRKKSFWLGMIALM-----LVFTMAFSASAAQPAKNV 36
QY 53 EYEYIVAF-NGYFTAKARNSFISSALKSSEVD-NWRIIPRNPSSDYPDFEVIQIKERQ 110
   | : | | : | : | : | : | : | : | : | : | : | : | : |
   37 EKDYIVGFKSGVKTASVKKDIKES--GGKVDKQFRIINAARAKLD----- 80
QY 111 KAGLLTLEDHNTKRVTPQKRVFSLKVAESDPTVPCNETRWSQKSSRLRRASLSLG 170
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
   81 KEALREKNDPDV-----ATVEED----- 99
QY 171 SGFWHATGRHSSRLRLRAIPROVAQTLDQADVLQMGYTCANVRVAVPTDGLSEKHPFKN 230
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
   100 -----HVAHALAQTPYGI-PLIKADKVAQCFKCANVAVLDTGIQASHDLNV 149
QY 231 VKERTNWTNERTLDDGLGHGTFVAGVTASMKRECO---GFAPDAELHIFRFTNNQVSYTS 287
   | : | | | | | | | | : | : | : | : | : | : | : | : |
   150 VGGASFVAGEAYNTDNGHGHVAGTVAALDNTTGVLGVAHSVLYAKVLNLSGSGSYTS 209
QY 288 WFLDAFNAILKIDVLNLSIGGDPFDND--HPFDKVKWELTANNVWVSAIGNDGPL--Y 343
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
   210 GLVSGIEWATTNGMDVNMISLGGSGSTAMKQAVDNAY---ARGVVVAAAGNSGSGWT 266
QY 344 GILNHPADQMDVIGVGIDFDENTARFSSRGMTTWELPGGYGRMKPDIITYAGVRGSGV 403
   | : | | | | : | : | : | : | : | : | : | : | : | : |
   267 NTIGYPKDYSTAVAGVDSNRSASFSGA---EL-----EVMAPGAGVYSTYP 314
QY 404 KGCRLSGTSVASPVVAGAVTLVSTVQKRELVPASMKQALIASARRLPVGNMFEQCH 463
   | : | | | | | | : | : | : | : | : | : | : | : | : |
   315 TTYATLNGTSMASPHVAGAAALILS---KHPNLSASQVRNRLSSTATYLG--SSPYTGK 369
QY 464 GKLDL 468
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
   370 GLINV 374

```

RESULT 11
Q45522
AC Q45522; PRELIMINARY; PRT; 382 AA.

```

DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Prepro-subtilisin Sendai precursor.
GN APRS.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G-825-6;
RX MEDLINE=95329264; PubMed=7605625;
RA Yamagata Y., Isshiki K., Ichishima E.;
RT "Subtilisin Sendai from alkalophilic Bacillus sp.: Molecular and
enzymatic properties of the enzyme and molecular cloning and
characterization of the gene. aprs.";
RL Enzyme Microb. Technol. 17:653-663(1995).
DR EMBL; D29698; BA006157.1; -.
DR HSSP; P29600; ICGI.
DR InterPro; IPR00209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Signal.
FT SIGNAL. 1 27 POTENTIAL.
FT CHAIN 114 382 SUBTILISIN SENDAI.
FT NON-TER 382 AA; 39901 MW; 9175AA1FD5B9DE43 CRC64;
SQ SEQUENCE 382 AA; 39901 MW; 9175AA1FD5B9DE43 CRC64;

Query Match 5.7%; Score 322; DB 2; Length 382;
Best Local Similarity 27.4%; Pred. No. 3.5e-15;
Matches 124; Conservative 59; Mismatches 172; Indels 98; Gaps 19;

QY 27 KKSFKAPCPGCSHLTKVFEFSIVVEYE-----YIVAFNGYFTAKARNSFISSALK-SS 80
   | | | | : | : | : | : | : | : | : | : | : | : | : | : |
   2 KKLFTKVVASAA--LLLSISLTATSVSAEOKKYLIGFENQLQV---TEFVSSDKGOS 56
QY 81 EVDNRIIPRNPSSDYPDFEVIQIKERKAGLLTLEDHNTKRVTPQKRVFSLKVAE 140
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
   57 ENSLFAEVNDESIEMLLEYEFEDIPVSVVE---LSPEDVKDLEK-DP-----SITVIE 105
QY 141 SD--PTVPCNETRWSQKSSRLRRASLSLGSFGFWHATGRHSSRLRLRAIPROVAQTLD 198
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
   106 EDIEVTITNQYTPW-----GTR-----VQ 125
QY 199 ADVLQMGYTCANVRVAVPTDGLSEKHPFKNKTERTNWT-NERTLDDGLGHGTFVAGVI 257
   | : | | | | | | | | : | : | : | : | : | : | : | : |
   126 APTATRTGITGVAVLDTGIS-THPDL-NIRGYSFVPGPSQDQNGHGHVAGTI 183
QY 258 ASMRE---CQGFAPDAELHIFRFTNNQVSYTSWFLDAFNAILKIDVLNLSIGGDPDM 314
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
   184 AALNNSIGVGVAVPAELYAKVLGANGSGSVSSIAOGLQWTAONNIHVANLSLGS----- 239
QY 315 DHPFDVKWELTAN-----NVIMVSAIGNDGLYGLTINNPADQMDVIGVGIDFDENTAR 369
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
   240 --PVGQSOTLEAVNQAATNAGVLVVAATNGNGS--GTVSYPARYANALAVGATDQNNRAS 295
QY 370 FSSRGMTTWELPGGYGRMKPDIITYAGVRGSGVKGKCRALSGTSVASPVVAGAVTLVVS 429
   | : | | | | : | : | : | : | : | : | : | : | : | : |
   296 FSOYGTGL-----NIVAPCGVIGTQSTYPCNRYASLSGTSNATPHVAGVAALV-- 341
QY 430 TVQKRELVPASMKQALIASARRLPVGNMFEQCH 462
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
   342 -KQKNPSWNTQIRQLTSTATSLGNSNQFGSG 373

```

RESULT 12
Q45301
ID Q45301 PRELIMINARY; PRT; 379 AA.
AC Q45301;
DT 01-NOV-1996 (Tremblrel. 01, Created)

```
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Subtilisin Carlsberg precursor.
GN SUBC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=15413;
RA Pan F.M., Chen S.T., Chion S.H.;
RT "Sequence characterization of the precursor of one mutant subtilisin
RL from Bacillus licheniformis."
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X91262; CAA62668.1; -.
DR HSP; P00780; ICSE.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_HIS; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 30 POTENTIAL.
FT NON_TER 107 379 POTENTIAL.
SQ SEQUENCE 379 AA; 5E6EDCA483EE1D6 CRC64;

Query Match 5.7%; Score 321; DB 2; Length 379;
Best Local Similarity 26.4%; Pred. No. 4.1e-15;
Matches 119; Conservative 62; Mismatches 165; Indels 104; Gaps 16;

QY 41 LTLKVFSSPV-----VEYEIVAF-NGYFTAKARNFISALKSSVD-NWRIPR 90
D 17 LVTFMEFSDSASAAQPKNKVEKYFVCKSGVATASVKKDIKES--GKVDKQFLINA 74
QY 91 NNPSSYPSDFEVIQKEQKAGLLIEDHPNTRKVTQPKVFRSLKYAESDTPVPCNET 150
D 75 AKATLD-----KEALKEVKNPDV-----AYVEED----- 99
QY 151 RWSOKQSSRPLRRASISLGSFGFWHATGRHSSRLLRAIPROVAQTLOADVLMQGYTGA 210
D 100 -----HVAHALGQTVPGI-PLIKADKVAQGFKA 129
QY 211 NRVAVFDTGLSEKHPHFNKVKERTNWTNERTLDDGLGCTFVAGVIASMQEQ---GFA 267
D 130 NRVAVLDTGIQASHPDNLNVVGCASFVAGEAYNTDGNHGHVAGIVAALDNTTGLVA 189
QY 268 PDALHIFRFTNNQVSYTSWFLDAFNAYALKKIDVNLISIGGPDFMD--HPFVCKWEL 325
D 190 PSVSLFAVKVNLSSGSGSYGIVSGIETWATTNGMDVINNSLGGPSGSGSTAMQAVDNAY-- 247
QY 326 TANNVIMVSAIGNDP--LYGTLLNADQMDVIGGIDPEDNIARFSSRGMTTWELPG 383
D 248 -SGVVPVAAAGSGSGTNTTIGYPAKYDSVLAAGVDSNSNRASFSSVGA---EL--- 300
QY 384 YGRMKPDIVTYGVRGSGVKGCRCALSGTSVSPVAGAVTLVSTVQKRELNVNPSMK 443
D 301 -----EVMAPGAGVYSTPTNTYATLNGTSMASPHVAGAAALILS---KHPNLSASQVR 351
QY 444 QALIASARKLPVNMFEQGHGKLDLLRAYQ 473
D 352 NRLSSATVILG--SSFYKGLINVEAAQ 379

RESULT 13
Q93L66
ID Q93L66 PRELIMINARY; PRT; 275 AA.
AC Q93L66;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Nattokinase (Fragment).
```

```
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Li S.Y.;
RT "Cloning and Expression of Nattokinase Gene in E.coli and Pachia
RL pastoris."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314856; CAC41625.1; -.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PROSITE; PS00136; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Kinase.
FT NON_TER 1 275
FT NON_TER 275
SQ SEQUENCE 275 AA; 3A13CCB3ADABF600 CRC64;

Query Match 5.7%; Score 320.5; DB 2; Length 275;
Best Local Similarity 31.1%; Pred. No. 2.6e-15;
Matches 92; Conservative 54; Mismatches 119; Indels 31; Gaps 11;

QY 187 RAIPROVAQTLOADVLMQGYTGANVRVAVFDTGLSEKHPHFNKVKERTNWTNERT--LD 244
D 2 QSVYPYGISQ-IPALHSGQYTGSGNVKVAVIDSGIDSSHPDL-NVRGASFVPSETPYQ 59
QY 245 DGLGHGTFVAGVIASMQEQ---COGFAPDAELHIFRFTNNQVSYTSWFLDAFNAYALKKI 301
D 60 DGSSHGTHVAGTIAALNNSIGVLGAPASLYAVKVLDDSTGSGOYSWIINGIEWAISNM 119
QY 302 DVNLNLSIGGPDFMD--HPFVCKWELTANNVIMVSAIGNDP--LYGTLLNADQMDVIG 357
D 120 GVINMSLGGSGSGTALKTVDKA---VSSGIVVAAAAGNAGSSGSGSTVGYPAKYPSTIA 176
QY 358 VGGIDFEDNTARFSSRGMTTWELPGYGRMKPDIVTYGVRGSGVKGCRCALSGTSVAS 417
D 177 VGAVNSSNRASFSSAGS---EL-----DVMAFGVSIQSLPGTYGAYNGTSMAT 224
QY 418 PVVAGAVTLVSTVQKRELNVNPSMKOALIASARLPVNMFEQGHGKLDLLRAYQ 473
D 225 PHVAGAAALILS---KHPTWTNAQVRDLRESTATYLG--NSFYKGLINVEAAQ 275

RESULT 14
Q9F943
ID Q9F943 PRELIMINARY; PRT; 374 AA.
AC Q9F943;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Subtilisin precursor (EC 3.4.21.62) (Fragment).
GN SUBC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 6816;
RX MEDLINE=20561724; PubMed=11109488;
RA Evans K.L., Crowder J.S., Miller E.S.;
RT "Subtilisins of Bacillus spp. hydrolyze keratin and allow growth on
RT feathers."
RL Can. J. Microbiol. 46:1004-1011(2000).
DR EMBL; AF205189; AAC31026.1; -.
DR HSP; P00780; ICSE.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_SER; 1.
```



```
Db 2 MRKSFGLMLTAFM-----LVFTMAFSDSASAAQPAKNV 36
Qy 53 EYEIVAF-NGYFTAKARNPSISSALKSSEVD--NRIIPRNPPSSDYPSPDFEVIQKEKQ 110
Db 37 EKDYIVGFKSGVKTASVKKDIKES--GGKVDKQFRINIINAKAKLD-----80
Qy 111 KAGLLTLEDHPNKRVTTPQKRVFRSLKYAESDPTVPCNETRWOSQKWSRPLRRASLSLG 170
Db 81 KEALKEVKNPDV-----AYVEED-----99
Qy 171 SGFWHATGRSSRRLLRAIPROVAQTLOADVLMQWGYTGANVRVAVEDTGLSEKHPFKN 230
Db 100 -----HVAHALAQTPVPGI--PLIKADKRVQAGFGKANVKVAVLDGTGIQASHPDLNV 149
Qy 231 VKERTNNTNERTLDGGLGHGTFVAGVIASRECO---GFAPDAELHIFRVTNNQVSYTS 287
Db 150 VGGASFVAGEAYNTDGNHGHGTHVAGTVAALDNTTGVLGVPASVSLYAVKVLNSSGSGSYS 209
Qy 288 WFLDAFNAILKIDVNLNLSIGGPDFMD--HPFVDKYWELTANNVIMVSAIGNDGPL--Y 343
Db 210 GIVSGIEWATTNGMDVINMISLGGASGTAMKQAVDNAY---ARGVVVVAAGNSGSGNT 266
Qy 344 GTLNNPADQMDVIGVGGIDFEDNIARESSRGMTTWELPGYGRMKPDIVTYGAGVRSKV 403
Db 267 NTIGYPAKYDSVIAVAGVDSNSRASFSGA---EL-----EVMAPGAGVYSTYP 314
Qy 404 KGGCRALSGTSVSPVAVAGVTLVSTVQKRELNPASMKOALIASARRLPGVNMFEQGH 463
Db 315 TNYATLNGTSMSPHVAGAAALILS---KPNLSASQVRNLSSTATYLQ--SSFYVGK 369
Qy 464 GKDLLRAYO 473
Db 370 GLINVEAAQ 379
```

RESULT 17

```
Q9F941 ID Q9F941 PRELIMINARY; PRT; 374 AA.
AC Q9F941;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Subtilisin precursor (EC 3.4.21.62) (Fragment).
GN SUBC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_Taxid=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC12759;
RX MEDLINE=20561724; PubMed=11109489;
RA Evans K.L., Crowder J.S., Miller E.S.;
RT "Subtilisins of Bacillus spp. hydrolyze keratin and allow growth on
feathers."
RL Can. J. Microbiol. 46:1004-1011(2000).
DR EMBL; AF205191; AAG31028.1; -.
DR HSP; P00780; ICSE.
DR InterPro; IPR00209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; P00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Signal.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 >374 SUBTILISIN.
FT NON_TER 374 374
SQ SEQUENCE 374 AA; 38423 MW; 640142EB813D0E60 CRC64;
```

Query Match 5.68; Score 316.5; DB 2; Length 374;
Best Local Similarity 24.9%; Pred. No. 8.5e-15;

```
Matches 121; Conservative 65; Mismatches 170; Indels 129; Gaps 17;
Qy 1 MKLVNIWLLIIVLLCGKKHLGDRLEKKSFEKPCGCSHLTLKVEFSSTV-----V 52
Db 2 MRKSFGLMLTAFM-----LVFTMAFSDSASAAQPAKNV 36
Qy 53 EYEIVAF-NGYFTAKARNPSISSALKSSEVD--NRIIPRNPPSSDYPSPDFEVIQKEKQ 110
Db 37 EKDYIVGFKSGVKTASVKKDIKES--GGKVDKQFRINIINAKAKLD-----80
Qy 111 KAGLLTLEDHPNKRVTTPQKRVFRSLKYAESDPTVPCNETRWOSQKWSRPLRRASLSLG 170
Db 81 KEALKEVKNPDV-----AYVEED-----99
Qy 171 SGFWHATGRSSRRLLRAIPROVAQTLOADVLMQWGYTGANVRVAVEDTGLSEKHPFKN 230
Db 100 -----HVAHALAQTPVPGI--PLIKADKRVQAGFGKANVKVAVLDGTGIQASHPDLNV 149
Qy 231 VKERTNNTNERTLDGGLGHGTFVAGVIASRECO---GFAPDAELHIFRVTNNQVSYTS 287
Db 150 VGGASFVAGEAYNTDGNHGHGTHVAGTVAALDNTTGVLGVPASVSLYAVKVLNSSGSGSYS 209
Qy 288 WFLDAFNAILKIDVNLNLSIGGPDFMD--HPFVDKYWELTANNVIMVSAIGNDGPL--Y 343
Db 210 GIVSGIEWATTNGMDVINMISLGGASGTAMKQAVDNAY---ARGVVVVAAGNSGSGNT 266
Qy 344 GTLNNPADQMDVIGVGGIDFEDNIARESSRGMTTWELPGYGRMKPDIVTYGAGVRSKV 403
Db 267 NTIGYPAKYDSVIAVAGVDSNSRASFSGA---EL-----EVMAPGAGVYSTYP 314
Qy 404 KGGCRALSGTSVSPVAVAGVTLVSTVQKRELNPASMKOALIASARRLPGVNMFEQGH 463
Db 315 TNYATLNGTSMSPHVAGAAALILS---KPNLSASQVRNLSSTATYLQ--SSFYVGK 369
Qy 464 GKDLL 468
Db 370 GLINV 374

RESULT 18
Q93QZ0 ID Q93QZ0 PRELIMINARY; PRT; 312 AA.
AC Q93QZ0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ISPD.
GN ISPD.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_Taxid=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=630; TRANSPOSON-TN5398;
RX MEDLINE=20106991; PubMed=10639372;
RA Farrow K.A., Lytras D., Rood J.I.;
RT "The macrolide-lincosamide-streptogramin B resistance determinant from
Clostridium difficile 630 contains two erm(B) genes."
RL Antimicrob. Agents Chemother. 44:411-413(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=630; TRANSPOSON-TN5398;
RA Farrow K.A., Lytras D., Rood J.I.;
RT "Genomic Analysis of the Erythromycin Resistance Element Tn5398 from
Clostridium difficile."
RL Microbiology 0:0-0(2001).
DR EMBL; AF109075; AAK77652.1; -.
DR MEROPS; S08.030; -.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
```



```
Db 274 AQ 275
RESULT 23
Q9DF2 PRELIMINARY; PRT; 310 AA.
AC Q9DF2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Kera (Fragment).
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=837B;
RA Goldstein G., Wagner E.K., Lee Y.E., Masisak E.S., Oh J.;
RT "Nucleotide sequence of kera gene encoding a keratinolytic protease of
Bacillus licheniformis OMU 837B.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF282895; AAG00494.1;
DR HSP: P00780; ICSE.
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
FT NOW_TER
SQ SEQUENCE 310 AA; 31285 MW; 740B9816277876EA CRC64;

Query Match 5.5%; Score 308; DB 2; Length 310;
Best Local Similarity 31.2%; Pred. No. 2.6e-14;
Matches 94; Conservative 49; Mismatches 130; Indels 28; Gaps 9;

QY 180 HSRRLRAIPROVAQTLDVLMQMGYTGANVRVAVFDTGLSKHPHFKN-VKERTNNNTN 239
Db 31 HVAHALAQTPYGI-PLIKADKVOAQGFKANVKVAVLDTGQASHDPDLNVVGASFAVG 89
QY 240 ERTLDGGLGHGTAVAGVIASMRQ---GFAPDAELHIFRVFTNNVSYTSWFLDAFNYA 296
Db 90 EAYNTDGNHGTAVAGTVAALDNTTGLGVAPSVSLYAVKLVNSGSGSYGVSGIEMA 149
QY 297 ILKIDVNLISIGGPDFMD--HPFVDKWLNTANNVIMVSAIGNDGPL--YGTLNPNADQ 352
Db 150 TTNGMDVNLISGSGSSTANKQAVDNAY--ARGVVVVAAGNSGSGNTNTIGYPKY 206
QY 353 MDVIGVGIDFEDNIARFSSRGMTTWELPGYGRMKPDIVTYIGVGRSGYKGCRAISG 412
Db 207 DSVIANGVDSNRSFSSVGA--EL-----EVNAPGAGYSTPTPTATYATLNG 254
QY 413 TVASPVVAGAVTLVSTVOKRELVPNSMKQALIASARRLPGVNMFQGHGKLDLLRAY 472
Db 255 TWSAPHVAGAAALILS---RHPNLSAQVRNRLSTATYLG--SSFYKGLINVEAAA 309
QY 473 Q 473
Db 310 Q 310

RESULT 24
Q45621 PRELIMINARY; PRT; 323 AA.
AC Q45621;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE A intracellular serine protease.
GN ISPO.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
```

```
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NKS-21;
RA Yamagata Y., Ichishima E.;
RT "A new alkaline serine protease from alkalophilic bacillus sp.:
cloning, sequencing, and characterization of an intracellular
protease.";
RL Curr. Microbiol. 30:357-366(1995).
DR EMBL: D37921; BAA07142.1;
DR HSP: Q99405; 1MPT.
DR MEROPS: S08.030;
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Protease.
SQ SEQUENCE 323 AA; 34643 MW; 6D20F71A73FIAD96 CRC64;

Query Match 5.4%; Score 304.5; DB 2; Length 323;
Best Local Similarity 28.8%; Pred. No. 5.1e-14;
Matches 91; Conservative 58; Mismatches 104; Indels 63; Gaps 14;

QY 187 RAIPROVAQTLDVLMQMGYTGANVRVAVFDTGLSKHPHFKN-VKERTNNNTNRTLD- 244
Db 19 KVIPPGI-EMTEAPAVWEAGYKGGNTVAVLDTGCTTHIEFKDQIIDGRNFTDDNSDP 77
QY 245 ---DGLGHGTAVAGVIASMRQ---GFAPDAELHIFRVFTNNVSYTSWFLDAFN 294
Db 78 DNVEDSNGHGTHTVCGPVAA---CENDKGVIGTAPKAKLVKVLGGQYDTKWIEGVR 134
QY 295 YAI-----LKKIDVNLISIGGPDFMDHPFVDK-VWELTANNVIMVSAIGNDGPLYGTLN 347
Db 135 YAINVRGPNRVRVIMSLSGR--IDTPELQAIIKHAVAEDILVVCAAGNE---GDGN 188
QY 348 NFADQM-----DIVGVGIDFEDNIARFSSRGMTTWELPGYGRMKPDIVTYIGVGRG 400
Db 189 HDTDEYAPGAYPEVQVGSVNLEGEISRFNTNCAI-----DLVAPGEEIIS 236
QY 401 SGVKGCRALSGTSVSPVAVAGAVTLVSTVOK---RELVPNSMKQAL-----IASARR 452
Db 237 TYLNGYAVLSGTSMTATPHVSGAALLIEQVEKEFERKLEPEIFAQLIKHTVSLNFSRR 296
QY 453 LFGVNMFEQGHGKLDL 468
Db 297 -----AQSGLLKL 305

RESULT 25
Q9F7C2 PRELIMINARY; PRT; 310 AA.
AC Q9F7C2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Keratinolytic protease Kera (Fragment).
GN KERA.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OMU1432B;
RA Goldstein G., Wagner E.K., Lee Y.E., Masisak E.S., Oh J.;
RT "Nucleotide sequence of kera gene encoding a keratinolytic protease of
Bacillus licheniformis OMU 1432B.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF282895; AAG10033.1;
```

DR HSP; P00780; 1CSE.
DR MEROPS; S08.001; -.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
KW Protease.
FT NON_TER
SQ SEQUENCE 310 AA; 31271 MW; 420B981623EEAD83 CRC64;

Query Match 5.4%; Score 304; DB 2; Length 310;
Best Local Similarity 30.9%; Pred. No. 5.1e-14;
Matches 93; Conservative 49; Mismatches 131; Indels 28; Gaps 9;

QY 180 HSSRLRLRAIPROVAQTLOADVLWQMYTCANVRVAFDTGLSEKHPFKNVKERTNWTN 239
DB 31 HVAHALAQVYPIGI-PLIRADKYQAQGFKGANVKVAVLDGTGQASHPDLLNVVGGASFVAG 89

QY 240 ERTLDGDLGHGTFVAGVIASMRECQ---GFAPDAELHIFRVFTNNQVSYTSWFDAFNVA 296
DB 90 EAYNTDGNHGTHVAGTVAALDNTTGVLGVPSPVSLYAVKVLNSSGSGSYSGIVSGIEWA 149

QY 297 ILKKIDVNLISGGPDFMD--HPFYDKVWELTANNVIMVSAIGNDGPL--YGTINNPADQ 352
DB 150 TTNGMDVINNSLGGSGSTAMKQAVDNAY---ARGVVVAAAGNSGSGSNTNTIGYPAKY 206

QY 353 MDVIGVGGIDFEDNIARESSRGMTTWELPGYGRMKPDIVTYCAGYRGSGVKGCCRALSG 412
DB 207 DSVIAGVADSNRASFSSVGA---EL-----EVNAPGAGVYSTYPTNTYATLNG 254

QY 413 TSVASPVVAGAVTLLVSTVQKRELVPASMKQALIASARRLPGVNMFEQGHGKLDLLRAY 472
DB 255 TSMGSPHVAGAAAILLS---KHPNLSASQVRNRLSSTATYLG--SSFYTGKGLINVEAAA 309

QY 473 Q 473
DB 310 Q 310

Search completed: May 29, 2003, 13:38:23
Job time : 106 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:28:27 ; Search time 25 seconds
(without alignments)
1745.324 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNIVLLVLLCGKKH.....PRVKRQLMQVHPKTPSV 1052

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5617	100.0	1052	1 MSIP_HUMAN	Q14703 homo sapien
2	5487	97.7	1052	1 MSIP_CRIGR	Q922a8 cricetulus
3	5476	97.5	1052	1 MSIP_RAT	Q9wtz3 rattus norv
4	5427	96.6	1052	1 MSIP_MOUSE	Q9wtz2 mus musculu
5	340	6.1	381	1 SUBT_BACST	P04142 bacillus st
6	339	6.0	381	1 SUBT_BACSU	P04189 bacillus su
7	338.5	6.0	381	1 SUBN_BACNA	P35335 bacillus su
8	338.5	6.0	381	1 SUBT_BACSA	P00783 bacillus su
9	329.5	5.9	379	1 SUBT_BACLI	P00780 bacillus li
10	319.5	5.7	382	1 SUBT_BACAM	P00782 bacillus am
11	316.5	5.6	275	1 SUBT_BACPU	P07518 bacillus pu
12	308	5.5	274	1 SUBD_BACLI	P00781 bacillus li
13	299	5.3	326	1 ISP_PAEPO	P29139 paenibacill
14	291	5.2	319	1 ISPL_BACSU	P11018 bacillus su
15	285	5.1	1398	1 PLS_PYRFU	P72186 pyrococcus
16	284.5	5.1	321	1 ISP_BACCS	P29140 bacillus cl
17	283	5.0	380	1 ELYA_BACCS	P27693 bacillus al
18	282	5.0	380	1 ELYA_BACCS	P41362 bacillus cl
19	267.5	4.8	269	1 SUBS_BACLE	P29600 bacillus le
20	266	4.7	378	1 ELYA_BACSP	P29074 bacillus sp
21	262.5	4.7	269	1 PRYM_BACSP	Q99405 bacillus sp
22	262.5	4.7	682	1 NISP_LACLA	Q07596 lactococcus
23	259.5	4.6	269	1 SUBB_BACLE	P29599 lactococcus
24	249	4.4	806	1 SUBV_BACSU	P29141 bacillus su
25	243	4.3	645	1 SUBE_BACSU	P16396 bacillus su
26	241.5	4.3	420	1 SUBT_BACSU	P28842 bacillus sp
27	240	4.3	401	1 THES_BACSP	P45670 bacillus sp
28	230.5	4.1	361	1 ELYA_BACPD	Q41363 bacillus ha
29	225.5	4.0	1433	1 SUBF_BACSU	P16397 bacillus su
30	223	4.0	279	1 THET_BACSU	P04072 thermocactin
31	220.5	3.9	595	1 BPRX_THBCU	P42780 bacteroides
32	220.5	3.9	894	1 WPRB_BACSU	P54423 bacillus su
33	219	3.9	422	1 TKSU_PYRKO	P58502 pyrococcus

34 215.5 3.8 461 1 EPIP_STAEP
35 215.5 3.8 1374 1 YQ56_CABEL
36 195.5 3.5 513 1 AOLI_THEAQ
37 195 3.5 384 1 PRFK_TRIAL
38 194 3.5 404 1 SMPL_MAGPO
39 193 3.4 635 1 PRTE_YEAST
40 191 3.4 1249 1 TPP2_HUMAN
41 190.5 3.4 1262 1 TPP2_MOUSE
42 190 3.4 1249 1 TPP2_RAT
43 189.5 3.4 1167 1 SCAL_STRPY
44 187.5 3.3 409 1 ALP_TRIHA
45 184.5 3.3 478 1 YSP3_YEAST

ALIGNMENTS

RESULT 1
MSIP_HUMAN
ID MSIP_HUMAN STANDARD; PRT; 1052 AA.
AC Q14703.09UE57;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ~~Membrane-bound~~ transcription factor site-1 protease precursor
DE (EC 3.4.21.-) (Site-1 protease) (Subtilisin/kexin-isozyme-1) (SKI-1).
GN MBTPS1 OR SIP OR SKI1 OR KIAA0091.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=20401147; PubMed=10944850;
RA Nakajima T., Iwaki K., Kodama T., Inazawa J., Emi M.;
RT "Genomic structure and chromosomal mapping of the human site-1
RT protease (SIP) gene.";
RL J. Hum. Genet. 45:212-217(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=95308325; PubMed=7788527;
RA Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,
RA Tabata S., Ishikawa K.-I., Kawarabayashi Y., Kotani H., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. III.
RT The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL PWA Res. 27:37-43(1995).
RN [3]
RP SEQUENCE OF 735-1052 FROM N.A.
RC TISSUE=Testis;
RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION, PARTIAL SEQUENCE, AND MASS SPECTROMETRY.
RX MEDLINE=20112790; PubMed=10644685;
RA Toure B.B., Munzer J.S., Basak A., Benjannet S., Rochemont J.,
RA Lazure C., Chretien M., Seidah N.G.;
RT "Biosynthesis and enzymatic characterization of human SKI-1/SIP and
RT the processing of its inhibitory prosegment.";
RL J. Biol. Chem. 275:2349-2358(2000).
RN [5]
RP SUBCELLULAR LOCATION.
RX MEDLINE=99145548; PubMed=9990022;
RA Seidah N.G., Mowla S.J., Hamelin J., Mamarbachi A.M., Benjannet S.,
RA Toure B.B., Basak A., Munzer J.S., Marcinkiewicz J., Zhong M.,
RA Barale J.-C., Lazure C., Murphy R.A., Chretien M., Marcinkiewicz M.;
RT "Mammalian subtilisin/kexin isozyme SKI-1: A widely expressed
RT proprotein convertase with a unique cleavage specificity and cellular
RT localization.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:1321-1326(1999).
CC ~~FUNCTION: Catalyzes the first step in the proteolytic activation~~

ID MSIP_CRIGR STANDARD; PRT; 1052 AA.
AC Q92A8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane-bound transcription factor site-1 protease precursor
DE (EC 3.4.21.-) (Site-1 protease) (Subtilisin/kexin-isozyme-1) (SKI-1)
GN MBTPS1 OR SIP OR SKIL
OS Cricetus griseus (Chinese hamster)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF ASP-218; HIS-249 AND SER-414.
RC STRAIN=25-RA;
RX MEDLINE=99026600; PubMed=9809072;
RA Sakai J., Rawson R.B., Espenshade P.J., Cheng D., Seegmiller A.C.,
RA Goldstein J.L., Brown M.S.;
RT "Molecular identification of the sterol-regulated luminal protease
RT that cleaves SREBPs and controls lipid composition of animal cells.";
RL Mol. Cell 2:505-514(1998).
RN [2]
RP FUNCTION.
RX MEDLINE=21111045; PubMed=11163209;
RA Ye J., Rawson R.B., Komuro R., Chen X., Dave U.P., Prywes R.,
RA Brown M.S., Goldstein J.L.;
RT "ER stress induces cleavage of membrane-bound ATF6 by the same
RT proteases that process SREBP.";
RL Mol. Cell 6:1335-1344(2000).
CC FUNCTION: Catalyzes the first step in the proteolytic activation
CC of the sterol regulatory element-binding proteins (SREBPs).
CC Other known substrates are BDNF and ATF6.
CC -1- CATALYTIC ACTIVITY: Cleaves after hydrophobic or small residues,
CC provided that Arg or Lys is in position P4. Cleaves known
CC substrates after Arg-Ser-Val-Leu (SERBP-2), Arg-His-Leu-Leu
CC (ATF6), Arg-Gly-Leu-Thr (BDNF) and its own propeptide after Arg-
CC Arg-Leu-Leu.
CC -1- COFACTOR: Calcium-dependent (By similarity).
CC -1- ENZYME REGULATION: Inhibited by divalent copper and zinc ions, but
CC not by nickel or cobalt. Inhibited by its prosegment, but not
CC smaller fragments thereof (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein; endoplasmic
CC reticulum and Golgi. May sort to other organelles, including
CC lysosomal and/or endosomal compartments (By similarity).
CC -1- INDUCTION: Down-regulated by sterols.
CC -1- PTM: The 148 kDa zymogen is processed progressively into two
CC membrane-bound 120 and 106 kDa forms in the endoplasmic reticulum,
CC and late into a secreted 98 kDa form. The propeptide is
CC autocatalytically removed through an intramolecular cleavage after
CC Leu-186. Further cleavage generates 14, 10, and 8kDa
CC intermediates (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF078105; AAC78321.1;
DR HSSP: Q99405; 1MPT.
DR MEROPS: S08.063;
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASIN.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease; Lipid metabolism;

KW Cholesterol metabolism; Signal; Transmembrane; Endoplasmic reticulum;
KW Golgi stack; Zymogen; Autocatalytic cleavage; Glycoprotein; Calcium.
FT SIGNAL 1 17 POTENTIAL.
FT PROPEP 18 186 POTENTIAL.
FT CHAIN 187 1052 MEMBRANE-BOUND TRANSCRIPTION FACTOR
FT SITE-1 PROTEASE.
FT LUMENAL (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT SERINE PROTEASE (POTENTIAL).
FT ARG/LYS/PRO-RICH (BASIC).
FT CLEAVAGE (AUTO-) (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT N-LINKED (GLCNAC...) (POTENTIAL).
FT N-LINKED (GLCNAC...) (POTENTIAL).
FT N-LINKED (GLCNAC...) (POTENTIAL).
FT N-LINKED (GLCNAC...) (POTENTIAL).
FT N-LINKED (GLCNAC...) (POTENTIAL).
FT D-N: LOSS OF ACTIVITY.
FT H->P: LOSS OF ACTIVITY.
FT S->A: LOSS OF ACTIVITY.
SQ SEQUENCE 1052 AA; 117564 MW; D7DBAFE988B4C69C CRC64;
Query Match Score 5487; DB 1; Length 1052;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;
QY 1 MKLVNIWLLLVLLCGKHLDRLEKKSFEKAPCGCSHLTLKVEFSSTVVEYEVAF 60
DB 1 MKLVNIWLLLVLLCGKHLDRLEKKSFEKAPCGCSHLTLKVEFSSTVVEYEVAF 60
QY 61 NGYFTAKARNSFISSALKSEVDNWRILPNNSSDYPDFEVIQIEKAGALLTLEDH 120
DB 61 NGYFTAKARNSFISSALKSEVDNWRILPNNSSDYPDFEVIQIEKAGALLTLEDH 120
QY 121 PNKRVTPQKVFSLKFAESDPTVPCNETRWQSKWSSRPLRRASISLGSFGFWHATGRH 180
DB 121 PNKRVTPQKVFSLKFAESDPTVPCNETRWQSKWSSRPLRRASISLGSFGFWHATGRH 180
QY 181 SSRLRLRAIPRQVAQTLOADVLMQMGYTGANRVAVEDTGLSEKHPKFNKERTNWTNE 240
DB 181 SSRLRLRAIPRQVAQTLOADVLMQMGYTGANRVAVEDTGLSEKHPKFNKERTNWTNE 240
QY 241 RTLDDGLGHGTFVAGVIASMRCEGFPADAEHLHFRVFTNNQVSYTSWFLDAFNAILKK 300
DB 241 RTLDDGLGHGTFVAGVIASMRCEGFPADAEHLHFRVFTNNQVSYTSWFLDAFNAILKK 300
QY 301 IDVLNLSIGGPDFMDHPFVDKVMWELTANNVIMVSAIGNDGLYGLTANNPADQMDVIGVG 360
DB 301 IDVLNLSIGGPDFMDHPFVDKVMWELTANNVIMVSAIGNDGLYGLTANNPADQMDVIGVG 360
QY 361 IDVEDNIAFSSRGMTTWELPGGYGRMKPDIVTYGAGVGRGVGKGRALSGTSVSPVY 420
DB 361 IDVEDNIAFSSRGMTTWELPGGYGRMKPDIVTYGAGVGRGVGKGRALSGTSVSPVY 420
QY 421 AGAVTLVSVQRELNVNPMQALIASARLPVGMVNFQGHGKLDLLRAYQILNSYKP 480
DB 421 AGAVTLVSVQRELNVNPMQALIASARLPVGMVNFQGHGKLDLLRAYQILNSYKP 480
QY 481 QASLSPSYIDLTCPYMWPCYSPYIYGGMTVVNVVTLNMGVTVGRVDPKQDPYLPQ 540
DB 481 QASLSPSYIDLTCPYMWPCYSPYIYGGMTVVNVVTLNMGVTVGRVDPKQDPYLPQ 540
QY 541 NGDNIEVAFSYSSVLPWPSGYLAISISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
DB 541 NGDNIEVAFSYSSVLPWPSGYLAISISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
QY 601 QTSTVKLPKIKVKIIPTPPRSKRVLDQYHNLRTPPGYPFRDNLBMKNDPLDWDGHDHHTN 660
DB 601 HTSTVKLPKIKVKIIPTPPRSKRVLDQYHNLRTPPGYPFRDNLBMKNDPLDWDGHDHHTN 660

```

QY 661 FRDMYQHLRSMGVFVEVLGAPTCFDASQVGTLLAWDSEEEFFPEETAKLRDVGNSL 720
CC
CC
CC 661 FRDMYQHLRSMGVFVEVLGAPTCFDATQVGTLLAWDSEEEFFPEETAKLRDVGNSL 720
CC
CC 721 VIFSDWYNTSMRKVKFYDENTROWMPDGTGGANIPALNELLVWNNMGFSDGLYGEFTL 780
CC
CC 721 VIFSDWYNTSMRKVKFYDENTROWMPDGTGGANIPALNELLVWNNMGFSDGLYGEFTL 780
CC
CC 781 ANHDMYASGSGTAKFPEDGVVITQTFKDGQLEVLKQETAVVENVPIGLYQIAPBGGGR 840
CC
CC 781 ANHDMYASGSGTAKFPEDGVVITQTFKDGQLEVLKQETAVVENVPIGLYQIAPBGGGR 840
CC
CC 841 IVLYGDSNCLDDSHRQKDCFLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
CC
CC 841 IVLYGDSNCLDDSHRQKDCFLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
CC
CC 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKNHOKLLSIDLKV 960
CC
CC 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKNHOKLLSIDLKV 960
CC
CC 961 VLPNRSNRPOVRPLSPGSGGAWDIPGIMPGRYNOEVGORTIPVAFGLAMVLAFFVQV 1020
CC
CC 961 VLPNRSNRPOVRPLSPGSGGAWDIPGIMPGRYNOEVGORTIPVAFGLAMVLAFFVQV 1020
CC
CC 1021 INKASRPRKPRKRVPRPOLMQOHPKTPSV 1052
CC
CC 1021 ISKASRPRKPRKRVPRPOLMQOHPKTPSV 1052
CC

RESULT 3
MSIP_RAT STANDARD; PRT: 1052 AA.
AC 09MT23;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane-bound transcription factor site-1 protease precursor
DE (EC 3.4.21.-) (Site-1 protease) (Subtilisin/kexin isozyme-1) (SKI-1).
GN MBTPS1 OR SIP OR SKI1.
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND SUBCELLULAR LOCATION.
RC STRAIN-Sprague-Dawley; TISSUE-Adrenal gland;
RX MEDLINE=99145548; PubMed=9990022;
RA Seidah N.G., Mowla S.J., Hamelin J., Mamarbachi A.M., Benjannet S.,
RA Toure B.B., Basak A., Munzer J.S., Marcinkiewicz J., Zhong M.,
RA Barale J.-C., Lazure C., Murphy R.A., Chretien M., Marcinkiewicz M.;
RT "Mammalian subtilisin/kexin isozyme SKI-1: A widely expressed
RT proprotein convertase with a unique cleavage specificity and cellular
RT localization.";
RL Proc Natl Acad Sci U S A. 96:1321-1326(1999).
CC
CC -1- FUNCTION: Catalyzes the first step in the proteolytic activation
CC of the steroid regulatory element-binding proteins (SREBPs).
CC Other known substrates are BDNF and Atf6.
CC -1- CATALYTIC ACTIVITY: Cleaves after hydrophobic or small residues,
CC provided that Arg or Lys is in position P4. Cleaves known
CC substrates after Arg-Ser-Val-Leu (SERBP-2), Arg-His-Leu-Leu
CC (Atf6), Arg-Gly-Leu-Thr (BDNF) and its own propeptide after Arg-
CC Arg-Leu-Leu.
CC -1- COFACTOR: Calcium-dependent (By similarity).
CC -1- ENZYME REGULATION: Inhibited by divalent copper and zinc ions, but
CC not by nickel or cobalt. Inhibited by its prosegment, but not
CC smaller fragments thereof (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein; endoplasmic
CC reticulum and Golgi. May sort to other organelles, including
CC lysosomal and/or endosomal compartments (By similarity).
CC -1- TISSUE SPECIFICITY: Widely expressed. In adult rat, highly
CC expressed in anterior pituitary, thyroid and adrenal glands and in
CC liver. In 2-day old rat, detected in developing skin, striated
CC

```

```

CC
CC muscles, cardiac muscles, bones, teeth and internal organs. Highly
CC expressed in retina, cerebellum, pituitary, submaxillary, thyroid
CC and adrenal glands, molars, thymus, kidney and intestine.
CC -1- PTM: The 148 kDa zymogen is processed progressively into two
CC membrane-bound 120 and 106 kDa forms in the endoplasmic reticulum,
CC and late into a secreted 98 kDa form. The propeptide is
CC autocatalytically removed through an intramolecular cleavage after
CC Leu-186. Further cleavage generates 14, 10, and 8kDa
CC intermediates (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF094821; AAD27011.1;
CC HSP: Q99405; 1MPT.
CC DR MEROPS: S08_063.
CC DR InterPro: IPR00209; Peptidase_S8.
CC DR Pfam: PF00082; Peptidase_S8; 1.
CC DR PRINTS: PR00723; SUBTILISIN.
CC DR PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
CC DR PROSITE: PS00137; SUBTILASE_HIS; 1.
CC DR PROSITE: PS00138; SUBTILASE_SER; 1.
CC KW Hydrolase; Protease; Serine protease; Lipid metabolism;
CC Cholesterol metabolism; Signal; Transmembrane; Endoplasmic reticulum;
CC Golgi stack; Zymogen; Autocatalytic cleavage; Glycoprotein; Calcium.
CC FT SIGNAL 1 17 POTENTIAL.
CC FT PROPEP 18 186 POTENTIAL.
CC FT CHAIN 187 1052 MEMBRANE-BOUND TRANSCRIPTION FACTOR
CC FT SITE 187 999 LUMENAL (POTENTIAL).
CC FT TRANSMEM 1000 1022 POTENTIAL.
CC FT DOMAIN 1023 1052 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 218 414 SERINE PROTEASE.
CC FT DOMAIN 1023 1050 ARG/LYS/PRO-RICH (BASIC).
CC FT SITE 186 187 CLEAVAGE (AUTO-) (BY SIMILARITY).
CC FT ACT_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 249 249 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 414 414 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT CARBOHYD 148 148 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 236 236 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 305 305 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 728 728 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT CARBOHYD 939 939 N-LINKED (GLCNAC...) (POTENTIAL).
CC FT SEQUENCE 1052 AA; 117479 MW; ECE568F2B41869DB CRC64;
CC
CC Query Match 97.58; Score 5476; DB 1; Length 1052;
CC Best Local Similarity 96.78; Pred No. 0;
CC Matches 101; Conservative 20; Mismatches 15; Indels 0; Gaps 0;
CC
CC QY 1 MKLVNWLILLVLLCGKHLGDRLEKKSFEKAPCPGCSHLTLKVFSTVVEYEVAF 60
CC DB 1 MKLVNWLILLVLLCGKHLGDRLEKKSFEKAPCPGCSHLTLKVFSTVVEYEVAF 60
CC
CC QY 61 NYFTAKARNSTISSALKSEVDNWRIRPNRNPSSDPSPDFVFIQIKKQKAGLLTLEDH 120
CC DB 61 NYFTAKARNSTISSALKSEVDNWRIRPNRNPSSDPSPDFVFIQIKKQKAGLLTLEDH 120
CC
CC QY 121 PNKRVTPQRFVPSLKFAESDPTVPCNETRMSQKWSRRPLRSLSGSGFWHATGRH 180
CC DB 121 PNKRVTPQRFVPSLKFAESDPTVPCNETRMSQKWSRRPLRSLSGSGFWHATGRH 180
CC
CC QY 181 SSRRLRLRAIPRQAQTQLQADVLWQMGYTGANVRVAFDTGLSEKHPHFKNKERTNWTNE 240
CC DB 181 SSRRLRLRAIPRQAQTQLQADVLWQMGYTGANVRVAFDTGLSEKHPHFKNKERTNWTNE 240
CC
CC QY 241 RTLDDGLGHGTFVAGVIASNRECOGFAPDAELHIFRVFTNNQVSYTSWFLDAFYAILKK 300

```

Db 241 RTLDGSGHGFVAGVATSMRECGFADPAELHFRFTNNQVSYTWFDAFYAIKK 300
Qy 301 IDVLNLSTGGDFDMDFVVKWELTANNVIMWSAIGNDGLYGTLANPADQMDVIGVG 360
Db 301 MDVLNLSTGGDFDMDFVVKWELTANNVIMWSAIGNDGLYGTLANPADQMDVIGVG 360
Qy 361 IDFDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVGRGKGRALSGTSSVSPV 420
Db 361 IDFDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVGRGKGRALSGTSSVSPV 420
Qy 421 AGAVTLVSTVQKRELNPASMKQALIASARLLPGVNMFGHCKLDDLRLRAYQILNSYK 480
Db 421 AGAVTLVSTVQKRELNPASVKQALIASARLLPGVNMFGHCKLDDLRLRAYQILNSYK 480
Qy 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMPTVNVNVLNGMGTGRIVDKPDQWPLPQ 540
Db 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMPTVNVNVLNGMGTGRIVDKPDQWPLPQ 540
Qy 541 NGDNIEVAFSYSSVLPWMSGYLAISISVTKKAASWEGIAQGHVIMITVASPAETESKNGAE 600
Db 541 NGDNIEVAFSYSSVLPWMSGYLAISISVTKKAASWEGIAQGHVIMITVASPAETESKNGAE 600
Qy 601 QTSVTKLPIKVIPTPPRSKRVLDQYHNLRYPPGYPFRDNLRMKNNDPLDWDGHDHITN 660
Db 601 HTSTVKLPIKVIPTPPRSKRVLDQYHNLRYPPGYPFRDNLRMKNNDPLDWDGHDHITN 660
Qy 661 FRDMYQHLRSMGVFVEVLGAPFCFDASQYGTLLMVDSEEEYPEETAKLRDQVNDGLSL 720
Db 661 FRDMYQHLRSMGVFVEVLGAPFCFDATQYGTLLMVDSEEEYPEETAKLRDQVNDGLSL 720
Qy 721 VIFSDWNTSVMRKVFYDENTQWMPDTPGGANIPALNELLSVWNNMGFSGLYEGEFTL 780
Db 721 VIFSDWNTSVMRKVFYDENTQWMPDTPGGANIPALNELLSVWNNMGFSGLYEGEFTL 780
Qy 781 ANHDMYASCSGTAKEPEDGVITQTFKDGLEVLKQETAVENVNPIGLYQIYPAEGGR 840
Db 781 ANHDMYASCSGTAKEPEDGVITQTFKDGLEVLKQETAVENVNPIGLYQIYPAEGGR 840
Qy 841 IVLYGDSNCLDDSHRQKDFWLLDALQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPERM 900
Db 841 IVLYGDSNCLDDSHRQKDFWLLDALQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPERM 900
Qy 901 EGNHLHRYSKVLEAHGDKPRPLPACPLRLSWAKPOPLNETAPSNLWKHKLISIDLKV 960
Db 901 EGNHLHRYSKVLEAHGDKPRPLPACPLRLSWAKPOPLNETAPSNLWKHKLISIDLKV 960
Qy 961 VLPNFRSNRQVRLPSGESGANDIPGIMPGRYNQEVGOTIPVFAFLGAMVLAFFVQ 1020
Db 961 VLPNFRSNRQVRLPSGESGANDIPGIMPGRYNQEVGOTIPVFAFLGAMVLAFFVQ 1020
Qy 1021 INKAKSRPKRRKPRVKRQPLMQCVHPKTPSV 1052
Db 1021 ISKAKSRPKRRKPRVKRQPLMQCVHPKTPSV 1052

RESULT 4

MS1P_MOUSE ID MS1P_MOUSE STANDARD; PRT; 1052 AA.
AC O9MT22.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane-bound transcription factor site-1 protease precursor
DE (EC 3.4.21.-) (Site-1 protease) (Subtilisin/kexin isozyme-1) (SKI-1)
DE (Sterol-regulated luminal protease).
GN MRPSP1 OR SIP OR SKI1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC TISSUE=Pituitary;
RX MEDLINE=99145548; PubMed=9990022;
RA Seidah N.G., Mowla S.J., Hamelin J., Mamabachi A.M., Benjannet S.,
RA Toure B.B., Basak A., Munzer J.S., Marcinkiewicz J., Zhong M.,
RA Barale J.C., Izure C., Murphy R.A., Chretien M., Marcinkiewicz M.,
RT "Mammalian subtilisin/kexin isozyme SKI-1: A widely expressed
RT proprotein convertase with a unique cleavage specificity and cellular
RT localization.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:1321-1326(1999).
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Catalyzes the first step in the proteolytic activation
CC of the sterol regulatory element-binding proteins (SREBPs).
CC Other known substrates are BDNF and ATF6 (By similarity).
CC -1- CATALYTIC ACTIVITY: Cleaves after hydrophobic or small residues,
CC provided that Arg or Lys is in position P4. Cleaves known
CC substrates after Arg-Ser-Val-Leu (SERBP-2), Arg-His-Leu-Leu
CC (ATF6), Arg-Gly-Leu-Thr (BDNF) and its own propeptide after Arg-
CC Arg-Leu-Leu.
CC -1- COFACTOR: Calcium-dependent (By similarity).
CC -1- ENZYME REGULATION: Inhibited by divalent copper and zinc ions, but
CC not by nickel or cobalt. Inhibited by its prosegment, but not
CC smaller fragments thereof (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein; endoplasmic
CC reticulum and Golgi. May sort to other organelles, including
CC lysosomal and/or endosomal compartments (By similarity).
CC -1- PFM: The 148 kDa zymogen is processed progressively into two
CC membrane-bound 120 and 106 kDa forms in the endoplasmic reticulum,
CC and late into a secreted 98 kDa form. The propeptide is
CC autocatalytically removed through an intramolecular cleavage after
CC Leu-186. Further cleavage generates 14, 10, and 8kDa
CC intermediates (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF094820; A027010.1;
CC EMBL: BC011533; A011533.1;
CC HSP; Q99405; IMPT.
CC MEROPS: S08.063;
CC MGD; MGI:1927235; Mbtps1.
CC InterPro: IPR000209; Peptidase_S8.
CC Pfam: PF00082; Peptidase_S8; 1.
CC PRINTS: PR00723; SUBTILISIN.
CC PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
CC PROSITE: PS00137; SUBTILASE_HIS; 1.
CC PROSITE: PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Protease; Serine protease; Lipid metabolism;
CC Cholesterol metabolism; Signal; Transmembrane; Endoplasmic reticulum;
CC Golgi stack; Zymogen; Autocatalytic cleavage; Glycoprotein; Calcium.
CC SIGNAL 1 17 POTENTIAL.
CC PROPEP 18 186 POTENTIAL.
CC CHAIN 187 1052 MEMBRANE-BOUND TRANSCRIPTION FACTOR
CC SITE-1 PROTEASE.
CC LUMENAL (POTENTIAL).
CC POTENTIAL.
CC TRANSMEM 1000 1022 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 1023 1052 SERINE PROTEASE.
CC FT DOMAIN 218 414 ARG/LYS/PRO-RICH (BASIC).
CC FT DOMAIN 1023 1050 ARG/LYS/PRO-RICH (BASIC).
CC FT SITE 186 187 CLEAVAGE (AUTO-) (BY SIMILARITY).
CC FT ACT_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 249 249 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 414 414 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT CARBOHYD 148 148 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 236 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 305 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 515 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 728 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 939 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 469 MISSING (IN REF. 2).
 SQ SEQUENCE 1052 AA; 117456 MW; FBD021C18775F2A CRC64;

Query Match 96.6%; Score 5427; DB 1; Length 1052;
 Best Local Similarity 95.8%; Pred. No. 0;
 Matches 1008; Conservative 25; Mismatches 19; Indels 0; Gaps 0;

QY 1 MKLVNIIWLLVLLCGKHLGDLRLEKSEKAPCPSCSHLTLKVEFSSTVEYEVAF 60
 DB 1 MKLVNIIWLLVLLCGKHLGDLRLEKSEKAPCPSCSHLTLKVEFSSTVEYEVAF 60

QY 61 NGYFTAKARNFISSALKSEVDNRIIPRNPPSSDPSDFEVIQIEKQKAGLLTLEH 120
 DB 61 NGYFTAKARNFISSALKSEVDNRIIPRNPPSSDPSDFEVIQIEKQKAGLLTLEH 120

QY 121 PNIKRVTQKRVFSLKYAESDPTVPCNETFWOKWQSRPLRRASLSLGSFWHATGRH 180
 DB 121 PNIKRVTQKRVFSLKYAESDPTVPCNETFWOKWQSRPLRRASLSLGSFWHATGRH 180

QY 181 SRRLLRAIPROVAOTLQADVLQMGYTGANRVAVFDLGLSEKHPKKNKERTWINE 240
 DB 181 SRRLLRAIPROVAOTLQADVLQMGYTGANRVAVFDLGLSEKHPKKNKERTWINE 240

QY 241 RLDDGLGHGTFVAGVYASRECQGFADAEHLIFRVFTNNOVSTVFWDAFNAILKK 300
 DB 241 RLDDGLGHGTFVAGVYASRECQGFADAEHLIFRVFTNNOVSTVFWDAFNAILKK 300

QY 301 IDVNLISGGPDFMDHPFVDKVMELTANNVIMVSAIGNDGPLYGLTNNPQDMVDVGGV 360
 DB 301 MDVNLISGGPDFMDHPFVDKVMELTANNVIMVSAIGNDGPLYGLTNNPQDMVDVGGV 360

QY 361 IDFNENIARFSSRGMTWELPGYGRVAPDITVYAGVSGVGGCRALSGTSVSPV 420
 DB 361 IDFNENIARFSSRGMTWELPGYGRVAPDITVYAGVSGVGGCRALSGTSVSPV 420

QY 421 AGAVTLVSTVQKRELNVNPKMQLIASARLPVGNMFEQGHGKLDLLRAYQILNSYK 480
 DB 421 AGAVTLVSTVQKRELNVNPKMQLIASARLPVGNMFEQGHGKLDLLRAYQILNSYK 480

QY 481 QASLSPSVIDTECPYMWPCYQIYGGMTVYVNTILNGMGTGRIVDKPQWOLYLPQ 540
 DB 481 QASLSPSVIDTECPYMWPCYQIYGGMTVYVNTILNGMGTGRIVDKPQWOLYLPQ 540

QY 541 NGDTEVAFSYSSVLPWPSGLAISIVTKKASWEGIAQGHVMTVWASPAETESKGA 600
 DB 541 NGDTEVAFSYSSVLPWPSGLAISIVTKKASWEGIAQGHVMTVWASPAETESKGA 600

QY 601 QTSVKLPIKVIIPPPSKRVLMQDQHNLRYPGYPFPRDLNRMKNDDPLDNGDHIHTN 660
 DB 601 HTSTVKLPIKVIIPPPSKRVLMQDQHNLRYPGYPFPRDLNRMKNDDPLDNGDHIHTN 660

QY 661 FRDMYOHLSRNGYFEVLGAPTCFASQXGTLMLVDSSEYFPEETAKLRDNDVNLGL 720
 DB 661 FRDMYOHLSRNGYFEVLGAPTCFASQXGTLMLVDSSEYFPEETAKLRDNDVNLGL 720

QY 721 VIFSDWNTSVMRKVFYDENTFRQWMDTGGANIPALNELLVSNMGMFSDGLYECEFTL 780
 DB 721 VIFSDWNTSVMRKVFYDENTFRQWMDTGGANIPALNELLVSNMGMFSDGLYECEFTL 780

QY 781 ANHDMYASGCSIAKPPEDGVVITOTFKDQGLEVLKQETAVVENVPILGLYQIPASGG 840
 DB 781 ANHDMYASGCSIAKPPEDGVVITOTFKDQGLEVLKQETAVVENVPILGLYQIPASGG 840

QY 841 IVLYGDSNCLDSDHQKDCFWLLDQLYTSYGVTPPSLSHSGNQRPPSGAGSVTPERM 900
 DB 841 IVLYGDSNCLDSDHQKDCFWLLDQLYTSYGVTPPSLSHSGNQRPPSGAGSVTPERM 900

QY 901 EGNHLHRSKYLEAHLDGPKRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDRV 960

DB 901 EGNHLHRSKYLEAHLDGPKRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDRV 960

QY 961 VLPNFRSNRPQVRPLSPGSGAWDIPGGIMPGRYNOEVGQITPVFAFLGAMVYLAFFVYQ 1020

DB 961 VLPNFRSNRPQVRPLSPGSGAWDIPGGIMPGRYNOEVGQITPVFAFLGAMVYLAFFVYQ 1020

QY 1021 INKAKSRPRKRRKPRKPRQPMQOVHPKTPSV 1052

DB 1021 ISKAKSRPRKRRRPRKPRQPMQOVHPKTPSV 1052

RESULT 5

ID SUBT_BACST STANDARD; PRT; 381 AA.

AC P29142:

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Subtilisin J precursor (EC 3.4.21.62).

GN APRI

OS Bacillus stearothermophilus.

OC Bacteria; Firmicutes; Bacillales; Geobacillus.

OX NCBI_taxid=1422;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCIMB 10278 / KCTC 1823;

RX MEDLINE=92231938; PubMed=1567435;

RA "Molecular cloning of a subtilisin J gene from Bacillus stearothermophilus and its expression in Bacillus subtilis.";

RT Biochem. Biophys. Res. Commun. 184:277-282(1992).

RL

CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE, IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity for peptide bonds, and a preference for a large uncharged residue in pl. Hydrolyzes peptide amides.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN IS NOT NECESSARY FOR NORMAL SPOULATION.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M64743; AAA22247.1; -

DR PIR; JQ1487; JQ1487.

DR HSSP; P04189; LSCJ.

DR MEROPS; S08.035; -

DR InterPro; IPR000209; Peptidase_S8.

DR Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE ASP; 1.

DR PROSITE; PS00137; SUBTILASE HIS; 1.

DR PROSITE; PS00138; SUBTILASE SER; 1.

KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal.

FT SIGNAL 1 29 POTENTIAL.

FT PROPEP 3 106 POTENTIAL.

FT CHAIN 107 381 SUBTILISIN J.

FT ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).

SQ SEQUENCE 381 AA; 39495 MW; C7A596F7629087D5 CRC64;

Query Match 6.1%; Score 340; DB 1; Length 381;
 Best Local Similarity 25.1%; Pred. No. 6.3e-17;

STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
IS NOT NECESSARY FOR NORMAL SPOULATION.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; K01988; AAA22742.1; -
DR EMBL; Y14083; CAA74536.1; -
DR EMBL; Y29109; CAB12870.1; -
DR EMBL; K01443; AAA22814.1; -
DR EMBL; M16639; AAA22744.1; -
DR EMBL; M31060; AAA22246.1; -
DR EMBL; M19125; AAA22245.1; -
DR PIR; A00372; SUBSI.; -
DR PIR; JSCJ; 13-JAN-99.
DR MEROPS; S08.036; -
DR Subtilisin; BG10190; aprE.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal;
KW 3D-structure; Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 106 POTENTIAL.
FT CHAIN 107 381 SUBTILISIN E.
FT ACT_SITE 138 138 CHARGE RELAY SYSTEM.
FT ACT_SITE 170 170 CHARGE RELAY SYSTEM.
FT ACT_SITE 327 327 CHARGE RELAY SYSTEM.
FT CONFLICT 27 27 V -> A (IN REF. 1, 4 AND 5).
SQ SEQUENCE 381 AA; 39523 MW; B7D2B38562C087D5 CRC64;

Query Match 6.0%; Score 339; DB 1; Length 381;
Best Local Similarity 25.1%; Pred. No. 7.4e-17;
Matches 123; Conservative 71; Mismatches 169; Indels 128; Gaps 18;

QY 1 MKLVNILLVLLCGKHGLDRLEKSKPEKAPCGCSHLTKVPSIVV-----52
DB 1 MSKKLNLISLFLT-----LIFTWFSNMSVQAAGKST 35

QY 53 EYEIVANGVFTAKRNGFISALKSEVDNRIIPRNPSSDYPSPDFEVIQEKQA 112
DB 36 EKRYIVGF-----KQTSAMSARKKDV-----SEKG 63

QY 113 GLTLLE-DHPNIRKVTQPKVFRSLKYAESDTPVPCNETRWQSKWSSRLRSLGSLGS 171
DB 64 GKVKQKQFYVNAATAALDEKAVKELK---KDPVSA-----95

QY 172 GFWHATGRHSRRLLRAIPROVAQTLQADVLQMGYTGANVRVAVFDTGLSEKHPFKNV 231
DB 96 ---YVEEDHIAHEYAQSVPGYISQ- IKAPALHSQGYTGSNVKAVIDSGIDSSHPLD-NV 150

QY 232 KERTNWTNERT--LDGGLGHTGFVAGVIA\$MRE---CQGFAPDRELHIFRVFNQVST 286
DB 151 RGSASFVPSNETPNYQDSSSHGTHVAGTIALNNSIGVLGSPSASLXAVKVLDSGQY 210

QY 287 SNFLDAFNATLKIDVNLNLSIGGPDMD--HPFVDKWELTANNVIMVSAIGNDGP--L 342
DB 211 SWINGIEMALSNMMDVINNSLGGFTG\$TALKIVDKA---VSSGIVVAAAAGNESSGS 267

QY 343 YCTLNAPADQMDVIGGIDFEDNIAFSSRGMTTWELPGYGRMKPDIVTYAGVRSG 402
DB 268 T\$TVGYPAKPTSLIANGAVNSNQAFSSAGS---EL-----D\$WAPGVSIQ\$TL 315

QY 403 VKGCRALSGTSVASPWAGAVTL\$VTVOKREL\$VNP\$KQALIASARLPGVNNFEGQ 462

DB 316 PGCTGAYNGTSMATPHVAGAAALILS---K\$PTWNAQVRDRLESTATYLG--NSFYTG 370
QY 463 HGKLDLLRAYQ 473
DB 371 KGLINVOAAQ 381

RESULT 7

ID SUBN_BACNA STANDARD; PRT; 381 AA.

AC P35835;
DC 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE Subtilisin NAT precursor (EC 3.4.21.62).
GN APRN.
OS Bacillus subtilis var. natto.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86029;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC2-1;
RX MEDLINE=93113095; PubMed=1369081;
RA Nakamura T., Yamagata Y., Ichishima E.;
RT "Nucleotide sequence of the subtilisin NAT gene, aprN, of Bacillus
subtilis (natto).";
RL Biosci. Biotechnol. Biochem. 56:1869-1871(1992).
CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
for peptide bonds, and a preference for a large uncharged residue
in pI. Hydrolyzes peptide amides.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
IS NOT NECESSARY FOR NORMAL SPOULATION.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation.
The European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D25319; BAA04989.1; -
DR EMBL; S51909; AAC60424.1; -
DR PIR; JH0778; JH0778.
DR HSP; P07518; IMEE.
DR MEROPS; S08.044; -
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 106 POTENTIAL.
FT CHAIN 107 381 SUBTILISIN NAT.
FT ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 381 AA; 39507 MW; DAED4B16ED1BA092 CRC64;

Query Match 6.0%; Score 338.5; DB 1; Length 381;
Best Local Similarity 27.1%; Pred. No. 8e-17;
Matches 118; Conservative 65; Mismatches 156; Indels 97; Gaps 17;

QY 48 SSTVEYEVIVANGVFTAKRNGFISALKSEVDNRIIPRNPSSDYPSPDFEVIQIK 107

```
Db 33 SST--EKKYIVGF-----KQMSAMSSAKKDDVI----- 59
QY 108 EKQKAGLLTLE-DHPNKRVTQKRVFSLKYAESDPTVPCNETRWSQKWSRPLRRAS 166
Db 60 -SEKGGKVKQKFKYVNAATAATLDEKAVKELK---KDPVA----- 95
QY 167 LSLGSGFWHATGRHSRRLLRAIPROVAQTLDVLMQGYTCANVRVAVFTGLSEKHP 226
Db 96 -----YVEDHIAHEYAQSVPGYISQ- IKAPALHSOGYTGNSVKVAVIDSGIDSSHP 146
QY 227 HFKNKVERTNWTNERT--LDDGLGHGTFVAGVIASMR---COGFAPDAELHIFRVTNN 281
Db 147 DL-NVRGGASFVPSETNPYQDSSHGTHVAGTIAALNNSIGVLGVSPASLYAVKVLDT 205
QY 282 QVSYTSWFLDAFNAILKKIDVNLNLSIGGDFMD--HPFVDKVKWELTANNVIMVSAIGND 339
Db 206 GSGQSWIINGIEWAISNNMDVINMSLGGPSSGTALKTVVDKA---VSSGIVVAAAAGNE 262
QY 340 GP--LYGTLLNPADQMDVICGCGIDFEDNIARESSRGMTWELPGGGRMKPDIVTYGAG 397
Db 263 GSSGSSSTVGYPAKYPTSTIAGVAVNSNQRASFSSAGS---EL-----DVMAPGVS 310
QY 398 VRGSGVKGCRALSGTSSVAPVAVAGAVTLLVSTVQKRELNVNPSMKQALIASARRPLGVN 457
Db 311 IQSTLPGGTYGAYNGTSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--N 365
QY 458 MFEQGHGKLDLLRAYQ 473
Db 366 SFYFGKGLINVAQAQ 381

RESULT 8
SUBT_BACSA
ID SUBT_BACSA STANDARD; PRT; 381 AA.
AC 200783;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE Subtilisin amylosacchariticus precursor (EC 3.4.21.62).
GN APR.
OS Bacillus subtilis var. amylosacchariticus.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1483;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89008194; PubMed=3139650;
RA Yoshimoto T., Oyama H., Honda T., Tone H., Takeshita T.,
RA Kamyama T., Tsuru D.;
RT "Cloning and expression of subtilisin amylosacchariticus gene.";
RL J. Biochem. 103:1060-1065(1988).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=72266687; PubMed=4560201;
RA Markland F.S., Kurihara M., Smith E.L.;
RT "Subtilisin Amylosacchariticus. II. Isolation and sequence of the
RT tryptic and cyanogen bromide peptides.";
RL J. Biol. Chem. 247:5602-5618(1972).
RN [3]
RP SEQUENCE OF 107-381.
RX MEDLINE=72266688; PubMed=5055784;
RA Kurihara M., Markland F.S., Smith E.L.;
RT "Subtilisin Amylosacchariticus. 3. Isolation and sequence of the
RT chymotryptic peptides and the complete amino acid sequence.";
RL J. Biol. Chem. 247:5619-5631(1972).
CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
CC for peptide bonds, and a preference for a large uncharged residue
CC in pI. Hydrolyzes peptide amides.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
CC SPORULATION, AND MANY MUTATIONS WHICH BLOCK SPORULATION AT EARLY
```

```
CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
CC IS NOT NECESSARY FOR NORMAL SPORULATION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D00264; BAA00186.1; -
DR PIR: A00971; SUBSS.
DR PIR: A41448; A41448.
DR HSSP: P04189; LSCJ.
DR MEROPS: S08.042; -
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE.ASP; 1.
DR PROSITE: PS00137; SUBTILASE.HIS; 1.
DR PROSITE: PS00138; SUBTILASE.SER; 1.
KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT PROPEP 31 106 POTENTIAL.
FT CHAIN 107 381 SUBTILISIN AMYLOSACCHARITICUS.
FT ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CONFLICT 191 191 S -> A (IN REF. 2).
FT CONFLICT 365 365 N -> D (IN REF. 2).
SQ SEQUENCE 381 AA; 39467 MW; 2251BADE22B4824F CRC64;

Query Match 6.0%; Score 338.5; DB 1; Length 381;
Best Local Similarity 26.8%; Pred. No. 8e-17;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;

QY 48 SSTVVEYEVAFNGYFTAKARNSSISSALKSSSEVDNWRIPRNPSSDYPSDFEVIQK 107
Db 33 SST--EKKYIVGF-----KQMSAMSSAKKDDVI----- 59
QY 108 EKQKAGLLTLE-DHPNKRVTQKRVFSLKYAESDPTVPCNETRWSQKWSRPLRRAS 166
Db 60 -SEKGGKVKQKFKYVNAATAATLDEKAVKELK---KDPVA----- 95
QY 167 LSLGSGFWHATGRHSRRLLRAIPROVAQTLDVLMQGYTCANVRVAVFTGLSEKHP 226
Db 96 -----YVEDHIAHEYAQSVPGYISQ- IKAPALHSOGYTGNSVKVAVIDSGIDSSHP 146
QY 227 HFKNKVERTNWTNERT--LDDGLGHGTFVAGVIASMR---COGFAPDAELHIFRVTNN 281
Db 147 DL-NVRGGASFVPSETNPYQDSSHGTHVAGTIAALNNSIGVLGVSPASLYAVKVLDT 205
QY 282 QVSYTSWFLDAFNAILKKIDVNLNLSIGGDFMD--HPFVDKVKWELTANNVIMVSAIGND 339
Db 206 GSGQSWIINGIEWAISNNMDVINMSLGGPSSGTALKTVVDKA---VSSGIVVAAAAGNE 262
QY 340 GP--LYGTLLNPADQMDVICGCGIDFEDNIARESSRGMTWELPGGGRMKPDIVTYGAG 397
Db 263 GSSGSSSTVGYPAKYPTSTIAGVAVNSNQRASFSSAGS---EL-----DVMAPGVS 310
QY 398 VRGSGVKGCRALSGTSSVAPVAVAGAVTLLVSTVQKRELNVNPSMKQALIASARRPLGVN 457
Db 311 IQSTLPGGTYGAYNGTSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--N 365
QY 458 MFEQGHGKLDLLRAYQ 473
Db 366 SFYFGKGLINVAQAQ 381

RESULT 9
SUBT_BACLI
```



```
FT HELIX 374 377
FT TURN 378 378
SQ SEQUENCE 379 AA; 39908 MW; F19A6DC5761FB504 CRC64;

Query Match
Best Local Similarity 5.9%; Score 329.5; DB 1; Length 379;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;

QY 1 MKLVNTWLLLLVLLGCKHGLDRLEKSKFEKAPCPGCSHLTLKVFSTV-----V 52
DB 2 MRKSKFWGLMTAFM-----LVTMAFSDSASAAQPAKV 36
QY 53 EYEYIVAF-NGYFTAKARNFSSALKSSVD-NWRIIPRNPPSSDPSDFEVIQIKEQ 110
DB 37 EKDYIVGFKSGVKTASVKKDIKES--GGRVDRQFRIINAAKALD----- 80
QY 111 KAGILLEDPHNTKRVTPQRKFRSLKYAESDPTVPCNETRWKQWSSRPLRASLSLG 170
DB 81 KEALKEYKNDPDV-----AYVEED----- 99
QY 171 SGFWHATGRHSSRRLLRAIPROVAQTLOADVLMQMGYTGANRVAVFDTLGLSEKHPHFKN 230
DB 100 -----HVAHALAGTVPGI-PLIKADKVAQGEKGANVAVLDTGIGASHPDLNV 149
QY 231 VKERTNWTNERTLDGLGHGTFVAGVIASMQEQ---GFAPDAELHIFRVTNNQVSYS 287
DB 150 VGGASFVAGEAYNTDGNHGHGTHVAGTVAALDNTTGVLGVPASVSLYAKVNLSSGSGTYS 209
QY 288 WFLDAFNAYLKKIDVNLNLSIGGDFMD--HPFVQKVMELTANNVIMVSAIGNDGPL--Y 343
DB 210 GIVSGIEWATNGTMDVNMISLGPSPGSGTAMKQAVDNAY---ARGVVVAAAAGNSGSGNT 266
QY 344 GTLNNDQMDQMDVLGVGGIDFEDNIAFSSRGMTTWELPGCYGRMKPDIVYGGVRSV 403
DB 267 NTGYPKYSVTAAGVADNSNRASFSSVGA--EL-----EVMAPGAGVYSTYP 314
QY 404 KGCRAISGTSVSPVAVAGAVTLLVSTVQKRELNVPSMKOALIASARRLPGVNMFPQGH 463
DB 315 TSYATLNGTSMASPHVAGAAALILS---KHPNLSASOVNRNLSSTATYLG--SSFYGK 369
QY 464 GKLDLLRAYQ 473
DB 370 GLINVEAAQ 379

RESULT 10
SUBT_BACAM
ID SUBT_BACAM STANDARD; PRT; 382 AA.
AC P00782;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Subtilisin BPN' precursor (EC 3.4.21.62) (Subtilisin Novo) (Alkaline
DE protease).
GN APR.
OS Bacillus amyloliquefaciens.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1390;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=ATCC 23844;
RX MEDLINE=85006739; PubMed=6090391;
RA Vasantha N., Thompson L.D., Rhodes C., Banner C., Nagle J.,
RA Filipula D.;
RT "Genes for alkaline protease and neutral protease from Bacillus
RT amyloliquefaciens contain a large open reading frame between the
RT regions coding for signal sequence and mature protein.";
RL J Bacteriol. 159:811-819(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84069812; PubMed=6316278;
RA Wells J.A., Ferrari E., Henner D.J., Estell D.A., Chen E.Y.;
RT "Cloning, sequencing, and secretion of Bacillus amyloliquefaciens
```

```
RT subtilisin in Bacillus subtilis.";
RL Nucleic Acids Res. 11:7911-7925(1983).
RN [3]
RP SEQUENCE OF 108-382.
RX MEDLINE=68086682; PubMed=6065094;
RA Markland F.S., Smith E.L.;
RT "Subtilisin BPN. VII. Isolation of cyanogen bromide peptides and the
RT complete amino acid sequence.";
RL J. Biol. Chem. 242:5198-5211(1967).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=72035041; PubMed=4399039;
RA Alden R.A., Wright C.S., Kraut J.;
RT "A hydrogen-bond network at the active site of subtilisin BPN'.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:119-124(1970).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF COMPLEX WITH INHIBITOR.
RX MEDLINE=85033707; PubMed=6387152;
RA Hirono S., Akagawa H., Mitsui Y., Iitaka Y.;
RT "Crystal structure at 2.6-A resolution of the complex of subtilisin
RT BPN' with streptomyces subtilisin inhibitor.";
RL J. Mol. Biol. 178:389-413(1984).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF MUTANT.
RX MEDLINE=90057412; PubMed=2684274;
RA Pantoliano M.W., Whitlow M., Wood J.F., Dodd S.W., Hardman K.D.,
RA Rollence M.L., Bryan P.N.;
RT "Large increases in general stability for subtilisin BPN' through
RT incremental changes in the free energy of unfolding.";
RL Biochemistry 28:7205-7213(1989).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RA Gallagher T., Oliver J., Bott R., Betzel C., Gilliland G.L.;
RT "Subtilisin BPN' at 1.6-A resolution: analysis for discrete disorder
RT and comparison of crystal forms.";
RL Acta Crystallogr. D 52:1125-1135(1996).
RN [8]
RP ACTIVE SITE.
RX MEDLINE=69104413; PubMed=5249818;
RA Markland F.S., Shaw E., Smith E.L.;
RT "Identification of histidine 64 in the active site of subtilisin.";
RL Proc. Natl. Acad. Sci. U.S.A. 61:1440-1447(1968).
CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
CC for peptide bonds, and a preference for a large uncharged residue
CC in PL. Hydrolyzes peptide amides.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- BIOTECHNOLOGY: Used as a detergent protease. Sold under the name
CC Alcalase by Novozymes.
CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
CC IS NOT NECESSARY FOR NORMAL SPOULATION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02496; AAB05345.1; -
CC PIR; A00970; SUBSN.
DR PDB; 1S01; 15-OCT-90.
DR PDB; 1S02; 15-JAN-92.
DR PDB; 1SBH; 07-DEC-95.
DR PDB; 1SBI; 07-DEC-95.
DR PDB; 1SBN; 31-JAN-94.
DR PDB; 1SBT; 31-MAY-84.
DR PDB; 2SBT; 31-MAY-84.
```

DR PDB: 1SIB: 31-OCT-93.
 DR PDB: 2SIC: 15-APR-93.
 DR PDB: 2SIC: 31-JAN-94.
 DR PDB: 2SIC: 31-JAN-94.
 DR PDB: 2SNI: 15-JAN-93.
 DR PDB: 2SPB: 15-OCT-93.
 DR PDB: 2STI: 15-JUL-91.
 DR PDB: 1STW: 15-JUL-91.
 DR PDB: 1SUA: 14-JAN-98.
 DR PDB: 1SUB: 31-JAN-94.
 DR PDB: 1SUC: 31-JAN-94.
 DR PDB: 1SUD: 31-JAN-94.
 DR PDB: 1SUE: 14-OCT-98.
 DR PDB: 1SUK: 14-NOV-95.
 DR PDB: 1AK9: 12-NOV-97.
 DR PDB: 1AU9: 31-DEC-97.
 DR PDB: 1AQN: 14-JAN-98.
 DR PDB: 1YJA: 11-JUL-96.
 DR PDB: 1YJB: 11-JUL-96.
 DR PDB: 1YJC: 11-JUL-96.
 DR PDB: 1A20: 29-APR-98.
 DR MEROPS: S08.034: --
 DR InterPro: IPRO00209: Peptidase_S8.
 DR Pfam: PF00082: Peptidase_S8: 1.
 DR PRINTS: PR00723: SUBTILISIN.
 DR PROSITE: PS00136: SUBTILASE_ASP: 1.
 DR PROSITE: PS00137: SUBTILASE_HIS: 1.
 DR PROSITE: PS00138: SUBTILASE_SER: 1.
 KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal;
 3D-structure. 1 32
 FT SIGNAL 33 107
 FT PROPEP 33 107
 FT CHAIN 106 382
 FT ACT_SITE 139 139
 FT ACT_SITE 171 171
 FT ACT_SITE 328 328
 FT HELIX 113 117
 FT TURN 118 119
 FT TURN 120 125
 FT TURN 126 127
 FT TURN 131 132
 FT STRAND 134 139
 FT TURN 144 145
 FT TURN 147 148
 FT STRAND 151 156
 FT TURN 159 160
 FT TURN 164 165
 FT HELIX 171 180
 FT TURN 193 194
 FT STRAND 196 201
 FT TURN 205 206
 FT HELIX 211 223
 FT TURN 224 225
 FT STRAND 228 231
 FT STRAND 235 235
 FT HELIX 240 251
 FT TURN 252 253
 FT STRAND 255 259
 FT TURN 267 268
 FT STRAND 274 277
 FT TURN 275 277
 FT TURN 279 280
 FT STRAND 282 287
 FT TURN 289 290
 FT STRAND 293 293
 FT TURN 295 296
 FT TURN 301 302
 FT STRAND 305 308
 FT STRAND 312 316
 FT TURN 317 319
 FT STRAND 320 324
 FT HELIX 327 344
 FT TURN 346 347

FT HELIX 350 359
 FT TURN 360 360
 FT STRAND 362 362
 FT HELIX 367 370
 FT TURN 371 372
 FT STRAND 374 374
 FT HELIX 377 381
 SQ SEQUENCE 382 AA: 39181 MW: ED987DAFA37B8335 CRC64;
 Query Match 5.7%; Score 319.5; DB 1; Length 382;
 Best Local Similarity 31.7%; Pred. NO. 1.9e-15;
 Matches 98; Conservative 50; Mismatches 118; Indels 43; Gaps 13;
 QY 180 HSRRLRAIPROVAQTQADVLQWQYTGANRVAVFDTGLSEKHPFK-----NYKE 233
 DB 102 HVHAYAQSVPGYSQ-IPKALHSQGTGTSNVKVAIVDSIDSSHPDLKLVAGGASMPVS 160
 QY 234 RTNWTNERTLDDGLGHGTFVAGVIASMR--CQGFAPDAELHIFRVFNNOVSYTSWFL 290
 DB 161 ETN-----PFQDNNSHGTHVAGTVAALNNSIGVLGVAPSASLYAVKVLGADGSGYSWII 215
 QY 291 DAFNVAILKIDVLNLSIGGDPMD--HPVDKVMWELTANNVIMVSAIGNDGP--LYGTL 346
 DB 216 NGIEWAIANNMDVINMSLGGPGSAAKAAVDKA---VAGVVVVAAGNEGTSQSSTV 272
 QY 347 NNPQDMQDVIGVGIGDEFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYCAGVRS--GVK 404
 DB 273 GYPGKYPVIAVGAVDSSNQRAFSSVGP---EL-----DVMAPGVSIQSTLPGNK 320
 QY 405 GCRALSGTSVASPVVAGAVTLVSTVQKRELVPNPKMKAALIASARRLPGVNMFEQHG 464
 DB 321 YG--AYNGTSMASPHVAGAAALILS---KHPNWTNVTQSRSLNTTTLG--DSFYKKG 373
 QY 465 KDLRLAYQ 473
 DB 374 LINVQAAQ 382
 RESULT 11
 SUBT_BACPU
 ID SUBT_BACPU STANDARD: PRT: 275 AA.
 AC P07518;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Subtilisin (EC 3.4.21.62) (Alkaline mesentericopeptidase).
 GN APR.
 OS Bacillus pumilus (Bacillus mesentericus).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1408;
 RN [1]
 RP SEQUENCE.
 RA Svendsen I., Genov N., Idakieva K.;
 RT "Complete amino acid sequence of alkaline mesentericopeptidase: a
 RT subtilisin isolated from a strain of Bacillus mesentericus.";
 RL FEBS Lett. 196:228-232(1986).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=92172311; PubMed=1793542;
 RA Dauter Z., Betzel C., Genov N., Papon N., Wilson K.S.;
 RT "Complex between the subtilisin from a mesophilic bacterium and the
 RT leech inhibitor eglin-C.";
 RL Acta Crystallogr. B 47:707-730(1991).
 CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
 CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 CC for peptide bonds, and a preference for a large uncharged residue
 CC in PL. Hydrolyzes peptide amides.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
 CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
 CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
 CC IS NOT NECESSARY FOR NORMAL SPOULATION.

FT	CHAIN	150	1398	PYROLYSIN.		
FT	ACT_SITE	179		CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT	ACT_SITE	365		CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT	ACT_SITE	441		CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT	CARBOHYD	152		N-LINKED (PROBABLE).		
FT	CARBOHYD	222		N-LINKED (POTENTIAL).		
FT	CARBOHYD	228		N-LINKED (POTENTIAL).		
FT	CARBOHYD	240		N-LINKED (POTENTIAL).		
FT	CARBOHYD	257		N-LINKED (POTENTIAL).		
FT	CARBOHYD	262		N-LINKED (POTENTIAL).		
FT	CARBOHYD	298		N-LINKED (POTENTIAL).		
FT	CARBOHYD	327		N-LINKED (POTENTIAL).		
FT	CARBOHYD	406		N-LINKED (POTENTIAL).		
FT	CARBOHYD	651		N-LINKED (POTENTIAL).		
FT	CARBOHYD	663		N-LINKED (POTENTIAL).		
FT	CARBOHYD	739		N-LINKED (POTENTIAL).		
FT	CARBOHYD	792		N-LINKED (POTENTIAL).		
FT	CARBOHYD	893		N-LINKED (POTENTIAL).		
FT	CARBOHYD	908		N-LINKED (POTENTIAL).		
FT	CARBOHYD	917		N-LINKED (POTENTIAL).		
FT	CARBOHYD	929		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1048		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1056		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1084		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1117		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1133		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1140		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1148		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1208		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1233		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1237		N-LINKED (POTENTIAL).		
FT	CARBOHYD	1332		N-LINKED (POTENTIAL).		
FT	CONFLICT	607		AKA -> PKP (IN REF. 1).		
FT	CONFLICT	881		Y -> H (IN REF. 1).		
FT	SEQUENCE	1398	AA; 154474 MW; 355D873A27D56552 CRC64;			
Query Match					5.1%; Score 285; DB 1; Length 1398;	
Best Local Similarity					20.5%; Pred. No. 4.2e-12;	
Matches 184; Conservative 115; Mismatches 316; Indels 284; Gaps 38;						
QY	245	DGLHGHTVAGVIA	-----	SMRCDQGFAPDAE	271	
DB	361	DGHGHTGTHVAGVYD	NDNDAMDWLSYSGEWEVSRLEYGMDYTNVTTDTVQGVAPGA	420		
QY	272	LHIFRVTNNQSVYTSW	LDAFNVAIKKIDVLNLSISGDPFMDHPFVD	-----	KVME 324	
DB	421	INAIKRLSDGRSMWD	LIEGMYAATHGADVISMISGG	-----	NAPYLDGTPDESVAUDE 476	
QY	325	LTAN-NWIVSAIGNDG	PLYGTLLNNPADQMDVIGV	-----	GIDFED	365
DB	477	LTEKYGVVFIAGNEG	PGINIVGSPVATKAITVGAAPVINGVYSQALGYPDYGF	536		
QY	366	-----	NIAFRSSGMMTWELPGYGRMKPDIVYAGVGRGSKV	-GGCRALSGTS	414	
DB	537	YFFPAYTNVRIAFSSR	-----	PRIDGEIKPNVAPAGVGIYSILPMWITGGADPFMSGTS	590	
QY	415	VASPVAGATYLLVSTV	KREL-VNPASKQALLASARLPG	-----	VNMEQGHKL 466	
DB	591	MAIPRSGVALLISGAR	EGIYINPDIIKVLGSGATWLEGDPTGQKTELDQGHGLV	650		
QY	467	DLLRAYQILNSKPOAS	LSUSPYIDLTECPYMWPCSOPIYGYGMPTVNVNITLNGMVG	526		
DB	651	NWTKSWELKA	-----	INGTLPIDVDHWADES	-----	YSDFAEYLGVDVITGLYARN 697
QY	527	RIVDPQWQPYLPQND	NIEVAFS-YSSVLM-PW-SGYLAISIS	-----	VTTKAASWEGIA 579	
DB	698	SIPDLVEM	-HKIYGDTEYRTEIAYTEPWIKPFSVGSIENNTPEVLVRKYDVEGLE	755		
QY	580	QGHVMTVASPATESK	NGAEOQTSTVKLPKIVKPIPTPPRSKRVLWDQ	-----	YHNL 631	
DB	756	PGLYVGRIIIDPPT	FPVIEDEILNTIVPEKF	-----	TPENNYTLTWYDINGPEMVFHF 811	
QY	632	RYPPG	-----	YFPR	-----	DNLRMKNDPLDWDGDIHTN 660

Db	812	TVPEGVYDVLVYAMTYDYGILYRDGMGFVFPYQDLYLPAAYSNPMPGNMELVWTG----	FN	867					
QY	661	FRDMYQHLSMGYFVEVILGAPFTC-----FDASQY-----GTLIMV	696						
Db	868	FAPLYE---SGFLRVYGVETIPSPVWVYINRTYLDNTWTFSEFINITNIIYAPINATLIPI	923						
QY	697	D-----SEEEYF-----PREIAKLRRVDN-----GLSLVIFSDWN-----	728						
Db	924	GLGTYNASVESVGDGEFFIKGIEVPEGTAEALKIRIGNSPVSPNSDLDLYDSKGNLVALD	983						
QY	729	-----TSVMRKVKFYDENTQWMP-----DTGGA-----	754						
Db	984	GNPATAEEVVVEYKPGVYSIVVHGYSVRDENGPNPTTTTFDLVVQMTLNGNKLKDSDI	1043						
QY	755	IPALNELLSV-----NMGFSGLYEGEFTLANIDMYASGCSIAKFPDGVVITQTF--	807						
Db	1044	ILGSNESVVVNTANITDRDHTPGTVSGIIEIRDNEVYQDNTSIKIPITLVIDKADFAV	1103						
QY	808	-----KDQGLEVLKQETAAVENPV-----ILGLYQIPABSGGRIVL-YGDSNCLDD	852						
Db	1104	GLTPABGVLGEARVYTLIVKHALTELPVNTAVIIGNYTYLTDENGTVFYAYTKLGD	1163						
QY	853	SHR---QKDCFWLLDALQ-----YYSYGVTPSPSLSHSGNRQRPSPSGASVTPPEWE	901						
Db	1164	EITVIVKRENTLEKFTQITVSEPEITDEENEFKLAMSS-----PENAAITVSVEME	1217						
RESULT 16									
ISP_BACCS									
ID	ISP_BACCS	STANDARD;	PRT;	321 AA.					
AD	P29140;								
DT	01-DEC-1992	(Rel. 24, Created)							
DT	01-DEC-1992	(Rel. 24, Last sequence update)							
DE	15-JUN-2002	(Rel. 41, Last annotation update)							
DE	Intracellular alkaline protease (EC 3.4.21.-).								
OS	Bacillus clausii.								
OS	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.								
NCBI_TaxID	=79880;								
OR	SEQUENCE FROM N.A.								
RP	STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;								
RC	Kato C., Nakano Y., Yamamoto M., Horikoshi K.;								
RL	Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.								
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.								
THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).									
EMBL;	D10730;	BA001573.1;							
PIR;	S27501;	S27501.							
RSP;	P00782;	LS01.							
MEKOPS;	S08.030;								
InterPro;	IPR000209;	Peptidase_S8.							
PrIMs;	PF00082;	Peptidase_S8;							
PRINTS;	PR00723;	SUBTILISIN.							
PROSITE;	PS00136;	SUBTILASE_ASP;							
PROSITE;	PS00137;	SUBTILASE_HIS;							
PROSITE;	PS00138;	SUBTILASE_SER;							
KW	Hydrolase; Serine protease.								
ACT SITE	49	49							
ACT SITE	86	86							
FT	ACT SITE	250							
FT	ACT SITE	250							
SEQUENCE	321 AA;	33747 MW;	621168D9F1044026	CR664;					
Query Match 5.1%; Score 284.5; DB 1; Length 321;									
Best Local Similarity 29.4%; Pred. No. 4.7e-13;									
Matches	95;	Conservative	49;	Mismatches 122;					
Matches	95;	Conservative	49;	Indels 57;					
Matches	95;	Conservative	49;	Gaps 14;					

QY 363 FEDNIARESSRGMTTWELPGYGRMKPDIVTYGAGVGRSGVGRGCRALSGTSSVAPWVAG 422
 DB 176 QNNRASESQYAGL-----DIVAPGVNQSTYPCGSTVASLNGTSMATPHVAG 223
 QY 423 AVTLVSVQKRELVPNPSMKQALIASARRLPQVGNFEOG 462
 DB 224 AAALV---KQKNPSWSNVQIRNLKNTATSLGSLNLYGSS 260

RESULT 24
 SUBV_BACSU
 ID SUBV_BACSU STANDARD; PRT; 806 AA.
 AC P29141, 1992 (rel. 24, Created)
 DT 01-DEC-1992 (rel. 24, last sequence update)
 DT 15-JUN-2002 (rel. 41, last annotation update)
 DE Minor extracellular protease vpr precursor (EC 3.4.21.-).
 GN VPR OR IPA-45R.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
 RX MEDLINE=92041574; PubMed=1938892;
 RA Sloma A., Rufo G.A. Jr., Theriault K.A., Dwyer M., Wilson S.W.,
 RA Pero J.;
 RT "Cloning and characterization of the gene for an additional
 RT extracellular serine protease of *Bacillus subtilis*.";
 RN J. Bacteriol. 173:6889-6895(1991).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
 RA Prescan E., Santanu M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.;
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees.";
 RL Mol. Microbiol. 10:371-384(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizon N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogawara A., Oudena B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale f.,
 RA Sato T., Scanlan E., Schleich S., Schroeter K., Scoffone F., Soldo B.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.F., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takanashi P., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viaroli A., Wambutt R., Wedler E., Wedler H., Yamane K., Yasumoto K., Yata K.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yoshikawa K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa K., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*

RT subtilis.";
 CC Nature 390:249-256(1997).
 CC -!- FUNCTION: NOT REQUIRED FOR GROWTH OR SPORULATION.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M76590; AAA22881.1; -;
 DR EMBL; X73124; CAA51601.1; -;
 DR EMBL; Z99123; CAB15835.1; -;
 DR PIR; A41341; A41341;
 DR HSSP; P00782; 2SBT.
 DR MEROPS; S08.UFA; -;
 DR Subtilisin; BG10591; vpr.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR002029; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 2.
 DR Pfam; PF02225; PA; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE.ASP; 1.
 DR PROSITE; PS00137; SUBTILASE.HIS; 1.
 DR PROSITE; PS00138; SUBTILASE.SER; 1.
 DR Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
 FT SIGNAL 1 28
 FT PROPEP 29 160
 FT CHAIN 161 806
 FT ACT_SITE 189 189
 FT ACT_SITE 233 233
 FT ACT_SITE 534 534
 FT SEQUENCE 806 AA; 8560 MW; F984E3BF0B869DDDD CRC64;
 SQ
 Query Match 4.48; Score 249; DB 1; Length 806;
 Best Local Similarity 20.2%; Pred. No. 7e-10;
 Matches 130; Conservative 69; Mismatches 166; Indels 278; Gaps 22;
 QY 25 LEKGFEPKAPCGCSHLTLKVEFSSTVVEYEVAFNGYFTAKARNSFISSALKSEVDN 84
 DB 62 LKEKSLAEKAGEAGSOSKSLKTART-----KAKNKAI-KAVKNGKYNR 104
 QY 85 WRIIPRNPPSDYDFEVIQIEKEKAGLLTLEDHPHNIKRVTPQKRVFSLKYAE---S 141
 DB 105 -----EYQVFGFSMK-----LPANEIPKLLAVKDVKAVYPNVYTKDNMK 146
 QY 142 DPTVPCNETRWSQKWSRPLRRASLSLGSFGFWHATGRHSRRLLRAIPRQVAQTLQADV 201
 DB 147 KDKVTISEDVSPQMDSDAPY-----IGAND 172
 QY 202 LQMCGYTCANVRVAVFDTLSEKHPKVKERTNWTNERTLD----- 244
 DB 173 ANDLGYTGKFKVAIDTGYEYNHPLKK-----NFGYKGYDFVNDYDKETPTGDP 227
 QY 245 -DGLGCHTFVAGVITASMRCECFADAEHLHFRVTNNQVYSWFLDAFNAILKKTVD 303
 DB 228 GEATDGHVAGTVAGNCTIKGVADATLLAYRVLPGSGSTENVIAGVERAVQGDV 287
 QY 304 LMLSTG-----GPDPMDFPVDFKRWELTANNVIMVSAIGNDGPLYGLTNNPADMDVIGV 359
 DB 288 MNLISLGNLNNPDWATSTALD--WAM-SEGVAVTSGNSGPNGWTVGSPGTSREASVG 344,
 QY 360 -----GI-----DFE----- 364
 DB 345 ATQPLPNEYAVTFGSSYSAKVMGYNKEDDVKALNNKEVELVEAGICEAKDFEGDLTKV 404
 QY 365 ----- 364

Db 405 AVVKGSGIAFVDKADNAKAGAIMGVVYNNLSGEIANVPCMSVPTIKLSLEDGEKLVSA 464
QY 365 -----DNIAFFSRG--MTTWELPGYGRMKPDIVTYGACV----- 398
Db 465 LKAGETKTTKLVSKALGQVADFSGRPMDTW-----MIKPDISAPGVNIVSTIP 517
QY 399 -----RSGVGKGCRCALSGTSVASPVVAGAVTLVSTVQKRELVPASMKQALIASAR 451
Db 518 THDPDHPYGYGSK-----QGTSMASPHIAGAVAVIKQAKPKVSV---EQIKAAIMTAV 568
QY 452 RLPGVN-----MFEQGHGKLDLLRAYOILNSYRQASLSP-SY 488
Db 569 TLKSDSGEVYPHNAQAG-----SARIMNAIKADSLVSPGSY 605

RESULT 25

SUBE_BACSU STANDARD; PRT; 645 AA.
AC P16396;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Minor extracellular protease EPR precursor (EC 3.4.21.-).
GN EPR OR IPA-15R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=168 / DE204;
RX MEDLINE=90340301; PubMed=2116590;
RA Brueckner R., Shoseyov O., Doi R.H.;
RT "Multiple active forms of a novel serine protease from Bacillus
subtilis";
RL Mol. Gen. Genet. 221:486-490(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89053875; PubMed=3142851;
RA Sloma A., Ally A., Ally D., Pero J.;
RT "Gene encoding a minor extracellular protease in Bacillus subtilis.";
RL J. Bacteriol. 170:5557-5563(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X53307; CAA37392.1; -
EMBL; M22407; AAA2423.1; -
EMBL; X73124; CAA51571.1; -
EMBL; Z99123; CAB15866.1; -
PIR; S11504; SUBSMP.
PIR; S39670; S39670.
HSP; O99405; IMPT.
MEROPS; S08.0PA; -
Subtilist; BG10561; epr.
InterPro; IPR000209; Peptidase_S8.

DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; zymogen; Signal; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 103 POTENTIAL.
FT CHAIN 104 645 MINOR EXTRACELLULAR PROTEASE EPR.
FT ACT_SITE 142 142 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 172 172 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 645 AA; 69695 MW; 401A4D5B60BE2E4A CRC64;

Query Match 4.38; Score 243; DB 1; Length 645;
Best Local Similarity 25.18; Pred. No. 1.3e-09;
Matches 104; Conservative 64; Mismatches 156; Indels 90; Gaps 19;

QY 99 SDFEVIQIKKQKAGLLTLED-----HPNIKRVTPQKVFSLKYAESDPTVPCN 148
Db 31 SEREVI-VVYKNKAGKETILSDADVEQYKHLPAVAVTADQETVKELK---QDPDILYV 86
QY 149 ETRWSQKWQSRPLRRASLSLGSFHWATGRHSRRLLRAIPROVAOTLQADVLQWQGYT 208
Db 87 ENNVSTAAADSTDFKVLK-----DGTDTSDNFEQWNLEPIQVQA-----WKAGLT 132
QY 209 GANVRVAVFDGTLSEKHPH-----FKNVKERTNWTNERTLDDGLGHGTFVAGVIASM 260
Db 133 GKNIKIAVIDSGIS---PHDDLSTAGYSAVSYTSSYKDDN-----GHGTHVAGIIGAK 183
QY 261 RE---COGFAPDAELHIFRVETNNQVSYTSWFLDAFNAILKKIDVLNLSTG--GPDFMD 315
Db 184 HNGVGIDGIAPAEQIYAVKALDQNGSGDLSLQGLIDWSIANRMDIVNMSLGTTSKIL 243
QY 316 HPFVQKVMWELTANNVIMVSAIGNDG---PLYGTINNADQMDYGVGGIDFEDNIARPS 372
Db 244 HDVANKAYE---QGVLLVAASGNDGNGKP---VNYPAAYSSVVAVASATNEKNOLASEST 296
QY 373 RGMTTWELPGYGRMKPDIVTY---GAGVRGSGVKGCRCALSGTSVASPVVAGAVTLVLS 429
Db 297 TG-----DEVEFSAPGNIISTYLNQYATGSGISQATPHRAAMFALL-- 339
QY 430 TVQKRELVPASMKQALIASARRLPGVNMFEEQGHGKLDLLRAYOILNSYKPKQAS 483
Db 340 --KQRD--PAETNVQLREEMRK----NIVDLGTAGRDQQQGYGLI-QYKAQAT 383

Search completed: May 29, 2003, 13:38:54
Job time : 30 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:33:32 ; Search time 50 Seconds
(without alignments)
2022.668 Million cell updates/sec

Title: us-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNIWLLLVLLCGKKH.....PRVKRQLMQVHPKTPSV 1052

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73.3

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5487	97.7	1052	T17093	intraluminal subti
2	1716	30.6	318	T43492	hypothetical prote
3	340	6.1	381	JQ1487	subtilisin (EC 3.4
4	338.5	6.0	381	1 SUBSS	subtilisin (EC 3.4
5	338.5	6.0	381	2 JH0778	subtilisin (EC 3.4
6	337.5	6.0	381	1 SUBSI	subtilisin (EC 3.4
7	329.5	5.9	379	1 SUBSCL	serine proteinase,
8	327.5	5.8	627	2 D75393	subtilisin (EC 3.4
9	322	5.7	382	2 I39780	subtilisin (EC 3.4
10	319.5	5.7	382	1 SUBSN	subtilisin (EC 3.4
11	312.5	5.6	272	2 A23624	subtilisin (EC 3.4
12	308	5.5	274	1 SUBSD	subtilisin (EC 3.4
13	304.5	5.4	323	2 I39867	microbial serine p
14	299	5.3	326	1 C41335	subtilisin (EC 3.4
15	295	5.3	275	2 JC1085	microbial serine p
16	294	5.2	321	2 JC5460	intracellular alka
17	293.5	5.2	322	2 G83922	intracellular alka
18	289	5.1	319	2 I39866	microbial serine p
19	285.5	5.1	374	2 I39781	subtilisin (EC 3.4
20	284.5	5.1	321	1 S27501	alkaline proteins
21	283	5.0	380	2 A49778	high-alkaline seri
22	281.5	5.0	1398	2 T28159	pyrolysin (EC 3.4
23	276.5	4.9	384	2 JC4802	alkaline proteins
24	270.5	4.8	328	2 G95392	probable protease
25	266	4.7	378	2 A33973	high-alkaline seri
26	260.5	4.6	682	2 S44131	subtilisin-like pr
27	259	4.6	757	2 C84120	subtilisin-type pr
28	257	4.6	1448	2 A12007	subtilase family p
29	253	4.5	397	2 JW0075	cysteine-dependent

alkaline serine pr
intracellular alka
surface layer-asso
microbial serine p
probable surface l
aerolysin precursor
serine proteinase
subtilisin (EC 3.4
probable serine pr
subtilisin (EC 3.4
subtilisin-type al
serine proteinase
intracellular alka
subtilisin-type al
proteinase [import
subtilisin-type pr

30 251.5 4.5 436 2 I39973
31 250 4.5 442 2 A69587
32 249.5 4.4 1345 2 T29090
33 249 4.4 806 2 A41341
34 246 4.4 1331 2 A72647
35 243.5 4.3 401 2 A57690
36 243 4.3 645 1 SUBSMP
37 242 4.3 419 1 S25835
38 241.5 4.3 412 2 T42024
39 241.5 4.3 420 1 S23407
40 241.5 4.3 792 2 H83736
41 240 4.3 401 2 I39974
42 239 4.3 444 2 B83891
43 236 4.2 372 2 D83735
44 235 4.2 488 2 A11930
45 231.5 4.1 799 2 G83753

ALIGNMENTS

RESULT 1

T17093

Intraluminal subtilisin-like proteinase SLP, membrane-bound - Chinese hamster
C:Species: Cricetus griseus (Chinese hamster)
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T17093

R:Sakai, J.; Rawson, R.B.; Espenshade, P.J.; Cheng, D.; Seegmiller, A.C.; Goldstein,

Molecular Cell 2, 505-514, 1998

A:Title: Molecular identification of the sterol-regulated luminal protease that cleav

A:Reference number: Z18677; MUID:99026600; PMID:9809072

A:Accession: T17093

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1052 <SA>

A:Cross-references: EMBL:AF078105; NID:g3892203; PID:g3892204; PIDN:AAC78321.1

A:Experimental source: strain 25-RA

C:Function:

A:Description: SLP is proposed to be a sterol-regulated protease that controls lipid

A:Note: SLP cleaves sterol regulatory element binding proteins (SREBPs) in the ER lum

Query Match 97.7%; Score 5487; DB 2; Length 1052;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKLVNIWLLLVLLCGKKHLDGRLEKKSFEKAPCCSHLTLLKVFSSVVEYEVAVF 60

DB 1 MKLVNIWLLLVLLCGKKHLDGRLEKKSFEKAPCCSHLTLLKVFSSVVEYEVAVF 60

QY 61 NGYFTAKARNSFTISSALKSSEVDNWRITIPRNPSSDYPSPDFEVIQIKKQKAGLLTLEH 120

DB 61 NGYFTAKARNSFTISSALKSSEVDNWRITIPRNPSSDYPSPDFEVIQIKKQKAGLLTLEH 120

QY 121 PNKRVTPORKVFRSLKYAESDPTVPCNETRWSQKQSSRPLRRASLSLGSFHWATGRH 180

DB 121 PNKRVTPORKVFRSLKYAESDPTVPCNETRWSQKQSSRPLRRASLSLGSFHWATGRH 180

QY 181 SSRRLRATPROVAQTLQADVLWQMGYTCANVRVAVFDGLSEKHPHFKNVKTWNTNE 240

DB 181 SSRRLRATPROVAQTLQADVLWQMGYTCANVRVAVFDGLSEKHPHFKNVKTWNTNE 240

QY 241 RTLDGGLGHGTFVAGVIASMRCCQGFAPDAELHIFRFTNNQVSYTSWFLDAFNAILKK 300

DB 241 RTLDGGLGHGTFVAGVIASMRCCQGFAPDAELHIFRFTNNQVSYTSWFLDAFNAILKK 300

QY 301 IDVLNLSIGGPDFMDHPFYDKYWEFLTANNVIMVSAIGNDGLYGLTNNPADQMDYGVGG 360

DB 301 IDVLNLSIGGPDFMDHPFYDKYWEFLTANNVIMVSAIGNDGLYGLTNNPADQMDYGVGG 360

QY 361 IDFDNIARFSSRGMTTWELPGYGRMKPDIVYAGVGRSGVKGCGCRALSSTSVASPVV 420

DB 361 IDFDNIARFSSRGMTTWELPGYGRMKPDIVYAGVGRSGVKGCGCRALSSTSVASPVV 420

QY 421 AGAVTLVSTVQKRELVPASMKQALIASARRLPVGNMFEGHKGKLDLRLAYQILNSYKP 480
 DB 421 AGAVTLVSTVQKRELVPASVKQALIASARRLPVGNMFEGHKGKLDLRLAYQILSSYKP 480
 QY 481 QASLSPSYDITLTCPTMYPYCSQPIYYGGMPTVNVYTLNGMGVTRGRVDPKPDWQPYLPQ 540
 DB 481 QASLSPSYDITLTCPTMYPYCSQPIYYGGMPTVNVYTLNGMGVTRGRVDPKPDWQPYLPQ 540
 QY 541 NGDNIEAFSYSSVLPWPSGYLAISIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 DB 541 NGDNIEAFSYSSVLPWPSGYLAISIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 QY 601 QSTVKLPKIKVIPTPPRSKRVLDQYHNLRYPPGYPRDLNLRKNDPLDNGDHHTN 660
 DB 601 HTSTVKLPKIKVIPTPPRSKRVLDQYHNLRYPPGYPRDLNLRKNDPLDNGDHHTN 660
 QY 661 FRDMYHLSRNGYFVEVLGAPFTCFDASQYGLTLLMVDSEYFPEETAKLRDNDVNGLSL 720
 DB 661 FRDMYHLSRNGYFVEVLGAPFTCFDASQYGLTLLMVDSEYFPEETAKLRDNDVNGLSL 720
 QY 721 VLFSDWNTSVMBKVFYDENTRQWMPDTGGANIPALNELLSSVWNGFSDGLYEGETL 780
 DB 721 VLFSDWNTSVMBKVFYDENTRQWMPDTGGANIPALNELLSSVWNGFSDGLYEGETL 780
 QY 781 ANHDMYASGCSIAKPPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEQGG 840
 DB 781 ANHDMYASGCSIAKPPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEQGG 840
 QY 841 IVLYGDSNCLDDSHRQKDFWLLDALLQYTSYGVTTPSLSHSGNRQPPSPGAGSVTPERM 900
 DB 841 IVLYGDSNCLDDSHRQKDFWLLDALLQYTSYGVTTPSLSHSGNRQPPSPGAGSVTPERM 900
 QY 901 EGNHLHRYSKVLAHGLDPKRPLPACPRLSNAKPOPLNETAPSLNWKHOKLLSIDLDKY 960
 DB 901 EGNHLHRYSKVLAHGLDPKRPLPACPRLSNAKPOPLNETAPSLNWKHOKLLSIDLDKY 960
 QY 961 VLPNFRSNRQVRLSPGSGAWDIPGGIMPGRYNOEGVQTIPIVFAFLGAMVLAFFVYQ 1020
 DB 961 VLPNFRSNRQVRLSPGSGAWDIPGGIMPGRYNOEGVQTIPIVFAFLGAMVLAFFVYQ 1020
 QY 1021 INKAKRPRKRRKRRVQKQVHPKTPSV 1052
 DB 1021 ISKAKRPRKRRRPRAKRPOLQQTTHPRTPSV 1052

RESULT 2
 T43492
 hypothetical protein DKFp434A219.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
 C:Accession: T43492
 R:Pouskka A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, December 1999
 A:Reference number: 222516
 A:Accession: T43492
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-318 <AAA>
 A:Cross-references: EMBL:AL133583
 A:Experimental source: adult testis; clone DKFp434A219
 C:Genetics:
 A:Note: DKFp434A219.1

Query Match 30.6%; Score 1716; DB 2; Length 318;
 Best Local Similarity 100.0%; Pred. No. 5.2e-115;
 Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 735 VKFYDENTRQWMPDTGGANIPALNELLSSVWNGFSDGLYEGETLANHDMYASGCSIA 794
 DB 1 VKFYDENTRQWMPDTGGANIPALNELLSSVWNGFSDGLYEGETLANHDMYASGCSIA 60
 QY 795 KPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEQGGRIYLVGDSNCLDDSH 854
 DB 795 KPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEQGGRIYLVGDSNCLDDSH 854

DB 61 KPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEQGGRIYLVGDSNCLDDSH 120
 QY 855 ROKDCFWLLDALLQYTSYGVTTPSLSHSGNRQPPSPGAGSVTPERMGNHLHRYSKVLEA 914
 DB 121 ROKDCFWLLDALLQYTSYGVTTPSLSHSGNRQPPSPGAGSVTPERMGNHLHRYSKVLEA 180
 QY 915 HLGDPKRPPLPACPRLSNAKPOPLNETAPSLNWKHOKLLSIDLDKYVLPNFRSNRQV 974
 DB 181 HLGDPKRPPLPACPRLSNAKPOPLNETAPSLNWKHOKLLSIDLDKYVLPNFRSNRQV 240
 QY 975 LSPGSGAWDIPGGIMPGRYNOEGVQTIPIVFAFLGAMVLAFFVQINKAKRPRKRRKPR 1034
 DB 241 LSPGSGAWDIPGGIMPGRYNOEGVQTIPIVFAFLGAMVLAFFVQINKAKRPRKRRKPR 300
 QY 1035 VRRPOLMOOVHPKTPSV 1052
 DB 301 VRRPOLMOOVHPKTPSV 318

RESULT 3
 JQ1487
 subtilisin (EC 3.4.21.62) J precursor - Bacillus stearothermophilus
 C:Species: Bacillus stearothermophilus
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 22-Jun-1999
 C:Accession: JQ1487
 R:Jang, J.S.; Kang, D.O.; Chun, M.J.; Byun, S.M.
 Biochem. Biophys. Res. Commun. 184, 277-282, 1992
 A:Title: Molecular cloning of a subtilisin J gene from Bacillus stearothermophilus an
 A:Reference number: JQ1487; MUID:92231938; PMID:1567435
 A:Accession: JQ1487
 A:Molecule type: DNA
 A:Residues: 1-381 <JNA>
 A:Cross-references: GB:M64743; NID:9142531; PIDN:AAA22247.1; PID:9142532
 A:Experimental source: strain NCIMB10278
 C:Genetics:
 A:Gene: aprJ
 A:Start codon: GTG
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: hydrolase; serine proteinase; zymogen
 F:1-29/Domain: signal sequence #status predicted <SIG>
 F:30-106/Domain: activation peptide #status predicted <ACT>
 F:107-381/Product: subtilisin J #status predicted <MAT>
 F:129-341/Domain: subtilisin homology <SBT>
 F:138,170,327/Active site: Asp, His, Ser #status predicted

Query Match 6.1%; Score 340; DB 2; Length 381;
 Best Local Similarity 25.1%; Pred. No. 2e-16;
 Matches 123; Conservative 71; Mismatches 169; Indels 128; Gaps 18;

QY 1 MKLYNIWLLVLLVLCGKKHLGDRLEKSKFEKAPCCSHLTKVEFSFVY----- 52
 DB 1 MKSKKMLISLLFALT-----LITAFSAMSQVQAGKSST 35
 QY 53 EYEYIVAFNGYTFKARNSTFSSALKSEVDNWRILRNPNPSSDYPDFEVIQKEKQKA 112
 DB 36 EKKYIVGF-----KQTSAMSASAKKDV-----SEKG 63
 QY 113 GLLTLE-DHPNIRKRVTPQRKVPFRSLKYAESDTPVPCNETRWSQKWSRRLRSLSGS 171
 DB 64 GKVKQKRYVNAATAATLDEKAVKELK-----KDPVA----- 95
 QY 172 GFWHATGRHSRRLRAIPRQVAQTLDQVLMQWGTGANVRVAVEDTGLSEKHPFKNV 231
 DB 96 ---VVEEDHIAHEVAQSVPYGISQ-IPALHSGQYTGSNVAVVIDGIDSSHPDL-NV 150
 QY 232 KERTNWTNERT--LDGLGCHGTFVAGVIASMR-----COGFAPDAELHIFRFTNNQVSYT 286
 DB 151 RCGASFVPESTNYQDSSSHGHVAGTIAALNNSIGVLGSPSASIAVKVLDSTGSGQY 210
 QY 287 SHFLDAFNALIKIDVNLNLSIGGDFMD--HPVDKRWELTANNVIMVSAIGNDGP--L 342
 DB 211 SWLINGEIAWISNMVDYNNKSLGGPSGSLKTVDKA---VSSGIVVAAAAGNAGSSGS 267

343	YGLTANNADQMDVTCVGGIDEFNIAFSRSGMTTHLP	PGCYGRMPDVTIVTGAVGRGSG	402
QY			
268	STVGYPKPYSTTRAVGAVNSNRQSFSGAGS---	EL-----DYNAPGVSIQSTL	315
Db			
403	VKGGCRALSGTSSVSPVVGAVTLLVSTVQKRELVP	SPASMKQALIASARLLPGVNNFEQG	462
QY			
316	PGGTGYGNGTSMATPHVAGAAALILS---	KHPTWTNAQVRDRLESTATYLG--NSFYFYG	370
Db			
463	HGKLDLLRAYQ	473	
QY			
371	KGLINVAQAAQ	381	
Db			

RESULT 4

SUBS

subtilisin (EC 3.4.21.62) amylosacchariticus precursor - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 24-Apr-1984 #sequence_revision 24-Feb-1995 #text_change 16-Jun-2000
C:Accession: A11448; A00971; S68013
R:Yoshimoto, T.; Oyama, H.; Honda, T.; Tone, H.; Takeshita, T.; Kamiyama, T.; Tsuru, D.
J. Biochem. 103, 1060-1065, 1988
A:Title: Cloning and expression of subtilisin amylosacchariticus gene.
A:Reference number: A11448; MUID:89008194; PMID:3139650
A:Accession: A11448
A:Molecule type: DNA
A:Residues: 1-381 <YOS>
A:Cross-references: GB:D00264; NID:g216328; PIDN:BA00186.1; PID:g912425
A:Experimental source: var. amylosacchariticus
R:Kurihara, M.; Markland, F.S.; Smith, E.L.
J. Biol. Chem. 247, 5619-5631, 1972
A:Title: Subtilisin amylosacchariticus. III. Isolation and sequence of the chymotryptic
A:Reference number: A00971; MUID:72266688; PMID:5055784
A:Accession: A00971
A:Molecule type: protein
A:Residues: 107-112-114; 148-152; 155-157; 164-170; 173-174; 178-181; 200-205; 210-212; 219-225;
A:Experimental source: var. amylosacchariticus
R:Kamal, M.; Hooeog, J.O.; Kaiser, R.; Shaifgat, J.; Razzaqi, T.; Zaidi, Z.H.; Joernvall, A.
FEBS Lett. 374, 363-366, 1995
A:Title: Isolation, characterization and structure of subtilisin from a thermostable Bac
A:Reference number: S68012; MUID:96069945; PMID:7589571
A:Accession: S68013
A>Status: preliminary
A:Molecule type: protein
A:Residues: 107-235; 'T', 237-245; 293-381 <KAM>
C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many
not necessary for normal sporulation.
C:Genetics:
A:Start codon: GTG
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:107-381/Product: subtilisin #status predicted <MAT>
F:129-341/Domain: subtilisin homology <SBT>
F:138-170.327/Active site: Asp, His, Ser #status predicted

Query Match	6.0%;	Score	338.5;	DB	1;	Length.381;			
Best Local Similarity	26.8%;	Pred. No.	2.6e-16;						
Matches	117;	Conservative	66;	Mismatches	156;	Indels	97;	Gaps	17;
QY	48	SSTVVEYEVIAFNGYFTAKARNSFFISSALKSSEVDNWRIIPRNPSSDPYDFEVIQIK	107			:			
Db	33	SSTI-EKKYVGF-----KQTSAMSSAKKKDVI-----	59						
QY	108	EKQKAGLLILE-DHPNIKRVTPOKRVFERSUKYAESDPTVPCNETRWSQKWSSRPLRRAS	166						
Db	60	-SEKGGVKQKQFYVNAATAATLDEKAVKELK-----KDPVA-----	95						
QY	167	LSLGSFHWATGRHSSRRRLRLRATPROVAOTLQADVLWQMGYTCGANRVVAVEDTCLSEKHP	226						
Db	96	-----VVEEDHIAHEVAQSVYPYGISQ-IKAPALHSQGYTGSNVKVAVIDSGIDSSHP	146						
QY	227	HFKNVKERTNWTNERT--LDDGLGHGFTFVAGYIASMRE---CGGFAPDAELHFRVFETNN	281						
Db	147	DL-NVRGGASFVPSFNYPQDSHGCTHVAGTTAAALNNSIGVLGVSPSASLVAVKVVDST	205						

	QY	282 QVSYTSWFLDAFNAYIAUKKIDVLNLSGPGPMD--HFFVDKWBELTANNVIMWSAIGND 339
		: : : : : : : : :
	Db	206 GSGOYSWIINGIEWAISNNDDVINMSLGGPSGSTALKTTVPDKA---VSSGIIVAAAAGNE 262
		: : : : : : : : :
	QY	340 GP--LYGLTNLPADQMDDVIGVGGDTEEDENIARFSSRGMTTWELPGGYGRMKRPDIITYGAG 397
		: : : : : : : : :
	Db	263 GSGGSSTGVGPYAKPYSTIAVGAVNSSNQRAFSFSAQS----EL-----DVNAPGVS 310
	QY	398 VSGSVGGCGRALSGTGSVASPVVACAVTLTVSTVQKRELVPASPMKQALIASARLLPGVN 457
		: : : : : : : : :
	Db	311 IQSTLPGGTGCAYNGTSMATPHVAGAAILILS---KHPTWTNAQVRDLRESTATYLGN-N 365
	QY	458 MFEQGKGKLDLLRAYQ 473
		: :
	Db	366 SFYYGKGLINVOQAAQ 381
		: :

RESULT 5

JH0778

subtilisin (EC 3.4.21.62) NAT precursor - *Bacillus subtilis* (strain natto NC2-1)
N;Alternate names: natto proteinase; nattokinase; subtilisin BSP
C;Species: *Bacillus subtilis*
C;date: 30-Sep-1993 #sequence.revision 30-Sep-1993 #text_change 21-Jul-2000
C;accession: JH0778; JS0601: JS0517: JC2036
R;Nakamura, T.; Yamagata, Y.; Ichishima, E.
Biosci. Biotechnol. Biochem. 56, 1869-1871, 1992
A;Title: Nucleotide sequence of the subtilisin NAT gene, aprN, of *Bacillus subtilis* {
A;Reference number: JH0778; MUID:93113095; PMID:1369081
A;Accession: JH0778
A;Molecule type: DNA
A;Residues: 1-381 <NAK>
A;Cross-references: GB:D25319; NID:g435439; PIDN:RAA04989.1; PID:g435440
R;Sumi, H.
Kagaku To Seibutsu 29, 119-123, 1991
A;Title: Natto kinase and fibrinolysis.
A;Reference number: JS0601
A;Accession: JS0601
A;Molecule type: protein
A;Residues: 107-381 <SUM>
R;Sumi, H.; Nakajima, N.
Nippon Nogelkagaku Kaishi 65, 1125-1127, 1991
A;Title: Studies on fibrinolysis enzymes in fermentation food.

```

Query Match          6.08; Score 338.5; DB 2; Length 381;
Best Local Similarity 27.1%; Pred. No. 2.6e-16;
Matches 118; Conservative 65; Mismatches 156; Indels 97; Gaps 17
QY 48 SSTVVEYETVAENGYETAKARNFISSAKSSVEDNWRRIIPNPSDDYDFEVIQIK 107
   |||  |  |||  |  |||  |  |||  |  |||  |  |||  |
Db 33 SST--EKKYITGF-----KQTSAMGSAKKKDDV----- 59

```

QY 108 EKQKAGLLTLE-DHPNKRVTQKRVFSLKYAESDPTVPCNETHKWSQKSSRPLRRAS 166
Db 60 -SEKGGKQKQFYVNAANAATLDERKAVKELK---KQPSVA----- 95
QY 167 LSGSGFWHATGRHSSRLRLRAIPROVAQTLDQVLMQGYTGANVRVAFDGLSEKHP 226
Db 96 -----YVEEDHIAHEYAQVYPYGISQ-IPKALPHSOGYTGSKNVKVAVIDSGIDSSHP 146
QY 227 HFKNKERTNNTERT--LDDGLGHGTGFAVGTASRE---CQGFAPDAELHIFRVFTNN 281
Db 147 DL-NVRGASFPVSETNPYQDGGSHGTHVACTIAALNNSIGVLGVAPSAIYAKVLDST 205
QY 282 QVSYTSWFLDAFNAYILKKIDVNLNLSIGGPDPMDD--HPFVQVKNWELTANNVIVWSAIGND 339
Db 206 GSGQYSWIINGIEAWISNNMVDVNNLSGGPTGTALKTVVDKA---YSSGIVVAAAAGNE 262
QY 340 GP--LYGTLPNADMDVIGVGIDFEDNTARSSGMMTWELPGCYGKMKPDIVTYGAG 397
Db 263 GSGSTSVIGPKYPTSTAVGAVSSNQRASTSSVGS---EL-----DYMAPGVS 310
QY 398 VRGSGVKGCRALSGTSSVAPVACAVTLVSVQKRELNVNPSKQMLTASARRLPGVN 457
Db 311 IQSTLPSTGYNGTSMAPHYAGAAALILS---KPTWTNAQVRDLSEIATYILG--N 365
QY 458 MFEQGHGKLDLLRAYQ 473
Db 366 SFYKGLGILNVAQAQ 381

RESULT 6
SUBSI
subtilisin (EC 3.4.21.62) E precursor - Bacillus subtilis
N:Alternate names: alkaline proteinase; bacillopeptidase E; extracellular alkaline serin
C:Species: Bacillus subtilis
C:Date: 04-Dec-1986 #sequence,revision 04-Dec-1986 #text_change 16-Jun-2000
C:Accession: A00972; A26116; I39970; I399778; I399779; S68012; H69586
R:Stahl, M.L.; Ferrari, E.
J. Bacteriol. 158, 411-418, 1984
A:Title: Replacement of the Bacillus subtilis subtilisin structural gene with an in vitro
A:Reference number: A00972; MUID:84212198; PMID:6427178
A:Accession: A00972
A:Molecule type: DNA
A:Residues: 1-381 <STA>
A:Cross-references: GB:K01988; NID:g143519; PIDN:AAA22742.1; PID:g143520
A:Experimental source: strain IL68
R:Wong, S.L.; Price, C.W.; Goldfarb, D.S.; Doi, R.H.
Proc. Natl. Acad. Sci. U.S.A. 81, 1184-1188, 1984
A:Title: The subtilisin E gene of Bacillus subtilis is transcribed from a sigma37 promoter
A:Reference number: A26116; MUID:84144862; PMID:6322190
A:Accession: A26116
A:Molecule type: DNA
A:Residues: 1-155 <WON>
A:Cross-references: GB:K01443; NID:g143665; PIDN:AAA22814.1; PID:g143666
R:Ikemura, H.; Takagi, H.; Inouye, M.
J. Biol. Chem. 262, 7859-7864, 1987
A:Title: Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*
A:Reference number: I39969; MUID:87222417; PMID:3108260
A:Accession: I39970
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-156 <K>
A:Cross-references: GB:M16639; NID:g143521; PIDN:AAA22744.1; PID:g143523
R:Henner, D.J.; Ferrari, E.; Perego, M.; Hoch, J.A.
J. Bacteriol. 170, 296-300, 1988
A:Title: Location of the targets of the hpr-97, sacQ36(Hy), and sacQ36(Hy) mutations in *Bacillus subtilis*
A:Reference number: I39778; MUID:88086885; PMID:2447063
A:Accession: I39778
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-8 <HEN>
A:Cross-references: GB:M19125; NID:g142527; PIDN:AAA22245.1; PID:g142528
R:Park, S.
J. Bacteriol. 171, 2657-2665, 1989

A:Title: Bacillus subtilis subtilisin gene (apre) is expressed from a sigma-A (sigma-A)
A:Reference number: I39779; MUID:89213955; PMID:2496113
A:Accession: I39779
A:Molecule type: DNA
A:Residues: 1-13 <PAR>
A:Cross-references: GB:M31060; NID:g142529; PIDN:AAA22246.1; PID:g142530
A:Experimental source: strain W168, substrate PY79
R:Kamal, M.; Hooe, J.O.; Kaiser, R.; Shafqat, J.; Razzaki, T.; Zaidi, Z.H.; Joernva
FEBS Lett. 374, 363-366, 1995
A:Title: Isolation, characterization and structure of subtilisin from a thermostable
A:Reference number: S68012; MUID:96069945; PMID:7589571
A:Accession: S68012
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 113-323 <KAM>
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.: Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabbret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.
Koetter, P.; Konningstein, G.; Koch, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y.M.; Ogawa, K.; Ogilwa, A.; Oudeghe, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Togononi, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zumbstein, A. Bacillus subtili
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: H69586
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-26, 'V', 28-381 <KUN>
A:Cross-references: GB:Z99109; GB:AL009126; NID:g2633260; PIDN:CAB12870.1; PID:g26333
A:Experimental source: strain 168
C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and m
not necessary for normal sporulation.
C:Genetics:
A:Gene: aprE
A:Map position: 690-771
A:Start codon: GTG
C:Function:
A:Description: catalyzes the hydrolysis of peptide bonds
A:Pathway: protein digestion
A:Note: this enzyme has broad specificity and will hydrolyze peptide amides, it prefe
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolyase; protein digestion; serine proteinase
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-106/Domain: activation peptide #status predicted <MPT>
F:107-381/Product: subtilisin E #status predicted <SBT>
F:129-341/Domain: subtilisin homology <SBT>
F:138,170,327/Active site: Asp, His, Ser #status predicted

Query Match Score 337.5; DB 1; Length 381;
Best Local Similarity 26.8%; Pred. No. 3e-16;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSTVVEYIVAFNGYETAKARNSFISSALKSEYDNWRIIPRNPSDDPSDFEVIQIK 107
Db 33 SST--EKYIVGF-----KQTSAMSSAKKDDVI----- 59
QY 108 EKQKAGLLTLE-DHPNKRVTQKRVFSLKYAESDPTVPCNETHKWSQKSSRPLRRAS 166
Db 60 -SEKGGKQKQFYVNAANAATLDERKAVKELK---KQPSVA----- 95
QY 167 LSGSGFWHATGRHSSRLRLRAIPROVAQTLDQVLMQGYTGANVRVAFDGLSEKHP 226
Db 96 -----YVEEDHIAHEYAQVYPYGISQ-IPKALPHSOGYTGSKNVKVAVIDSGIDSSHP 146
QY 227 HFKNKERTNNTERT--LDDGLGHGTGFAVGTASRE---CQGFAPDAELHIFRVFTNN 281

Db 147 DL-NVRGASFPVSETPNYQDSSGHGTHVAGTAAALNNSIGVLGVSPTSASLYAVKVLDT 205
QY 282 QVSTSMFLDAENVAIKKIDVLNLSIGGPDFMD--HPFVDKRWELTANNVIMVSAIGND 339
Db 206 GSQYSLWINGIEWASNNMDVNLMSLGGPTGTALTQVVDKA---VSSGIVVAAAAGNE 262
QY 340 GP--LYGTLLNPADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
Db 263 GSSGSTSTGVYPAKYPSTIAGVAVNSNQASFSAGS---EL-----DVMAPGVS 310
QY 398 VSGGVKGGCRALSGVSAPVAVGAVTLVSTVQRELVPASMKQALIASARRLPQVN 457
Db 311 IQSTLPGGTGAVNGTSMATPHVAGAAALILS---KPTWTNAQVRDRLESTATYLG--N 365
QY 458 MFEQGHGKLDLLRAYQ 473
Db 366 SFYIGKGLINVAQAQ 381

RESULT 7

SUBSCL
subtilisin (EC 3.4.21.62) Carlsberg precursor - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 21-Jul-2000
C:Accession: A24111; A00968
R:Jacobs, M.; Eliasson, M.; Uhlen, M.; Flock, J.I.
Nucleic Acids Res. 13, 8913-8926, 1985
A:Title: Cloning, sequencing and expression of subtilisin Carlsberg from Bacillus licheniformis
A:Reference number: A24111; MUID:86093688; PMID:3001553
A:Accession: A24111
A:Molecule type: DNA
A:Residues: 1-379 <JAC>
A:CROSS-references: GB:X03341; NID:G487721; PIDN:CAB56500.1; PID:G5921206
A:Experimental source: strain NCIB6816
R:Smith, E.L.; DeLange, R.J.; Evans, W.H.; Landon, M.; Markland, F.S.
J. Biol. Chem. 243, 2184-2191, 1968
A:Title: Subtilisin Carlsberg. V. The complete sequence; comparison with subtilisin BPN'
A:Reference number: A00968; MUID:68234702; PMID:4967581
A:Accession: A00968
A:Molecule type: protein
A:Residues: 106-206, 'S', 208-232, 'A', 234-261, 'N', 263-264, 'S', 266-315, 'N', 317-379 <SMT>
C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many not necessary for normal sporulation.
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-105/Domain: propeptide #status predicted <APT>
F:106-379/Product: subtilisin Carlsberg #status experimental <MPT>
F:128-339/Domain: subtilisin homology <SBT>
F:137,168,325/Active site: Asp, His, Ser #status predicted

Query Match 5.9%; Score 329.5; DB 1; Length 379;
Best Local Similarity 25.3%; Pred. No. 1.1e-15;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;
QY 1 MKLVNIWLLLVLLVCKHGLDRLEKKSEKAPCPGCSHLTLKVFSTV-----V 52
Db 2 MRKKSFWLGMATFM-----LVFTMAFSDSASAAQAKNV 36
QY 53 EYEYIYAP-NGFTAKARNSFTISSALKSEVD-NWRIIPRNPSSDYPSPDFEVIQKEQ 110
Db 37 EKDIYVGFSGKVTASVKKDIKES--GGKVDKQFRIINAARAKLD----- 80
QY 111 KAGLLLEDPHNPTRPQKRVSLKYAESDPTVPCNETRWSQKWSRPLRASLSLG 170
Db 81 KEALKEVKNDPDV-----AYVEED----- 99
QY 171 SGFWHATGRSRRLLRAIPROVAOTLQADVLWQMGYTGANVRVAVFDTLGSEKHPHEKN 230
Db 100 -----HVAHALAQTPYGI-PLIKADKVAQGFKANVKAVIDLGIOASHPDLVN 149
QY 231 VKERTNWNERTLDDGLGHTGVAGVIASMRQCG---GFAPDAELHIFRVFTNNQVSYTS 287

Db 150 VVGASFVAGEAYNTDGNHGHVAGTVAALDNTTGLVGVAPSVSLYAVKVLNSSGSGTYS 209
QY 288 WFLDAFNAYIAIKKIDVLNLSIGGPDFMD--HPFVDKRWELTANNVIMVSAIGNDGPL--Y 343
Db 210 GIYSGIEWATNGMDVINMSLGGPSGTAMQAVDNAY---ARGVVVVAAGNSGSGNT 266
QY 344 GTLNNPADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVG 403
Db 267 NTIGYPAKYDSVIAVGAVDNSNRASFSSVGA---EL-----EVMAPGAGVYTYTP 314
QY 404 KGCGRALSGVSAPVAVGAVTLVSTVQRELVPASMKQALIASARRLPQVNMPQGH 463
Db 315 TSTYATLNGTSMATPHVAGAAALILS---KHPNLSASQVRNRLSSTATYLG--SSFYIGK 369
QY 464 GKLDLLRAYQ 473
Db 370 GLINVEAAAQ 379

RESULT 8

D75393
serine proteinase, subtilase family - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: D75393
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: D75393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-627 <WHI>
A:CROSS-references: GB:AE001990; GB:AE000513; NID:G6459214; PIDN:AAF11026.1; PID:G645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRI459
A:Map position: 1

Query Match 5.8%; Score 327.5; DB 2; Length 627;
Best Local Similarity 27.5%; Pred. No. 3.5e-15;
Matches 119; Conservative 62; Mismatches 181; Indels 71; Gaps 16;

QY 57 IVAFNGVFTAKARNSF--ISSALKSEVDNRRIIPRNPSSDYPSPDFEVIQKEKAGL 114
Db 47 IVRFNVAQAQGRALFKNLRGQLNSQ-----IAKLGPSAGF-----LQAVNSOKATQ 94
QY 115 LTLEDHPNTRKRVTPQKRVSLKYAESDPTVPCNETRWSQKWSRPLRASLSLGSGFW 174
Db 95 LWLDQSVLLPMTVQARAVSOLPFVAA-----VFENFKVQIPRVVAMSNASAPAG-- 144
QY 175 HATGRHSRRLLRAIPROVAOTLQADVLWQMGYTGANVRVAVFDTLGSEKHP----- 226
Db 145 --TPTH-----LQIGAPAAWAAAGFGKGNIRIGHLDSDIDPSHPELAGKVA 189
QY 227 --HFNKVBKERTNWTNERTLDDGLGHTGVAGVIASMRQCGFAPDAELHIFRVFTNNQVS 284
Db 190 FQEFNGEGDVRSSQPHDITD-----HGHTAGLLVGSK--VGVAPGARVISALVLPNNEG 243
QY 285 YT-----SWFLDAFNVAIILK-IDVLNLSIGGPDFMDHPFDVKVWELTANNVIMVSAIG 337
Db 244 FAQVIAGMQVLDPDNADTDDGADVNNMSLGLPGTWN--FIYVNNMLKAGVVPVFAIG 302
QY 338 NDGPLYGTLLNPADQMDVIGVGIDFEDNIARFSSRGMTTW--ELPGGYGRMKPDIVTYG 395
Db 303 NFGPAAGSTGSPCNLPAIGVGAVDNSNQVASSPSRGPVAVWQGEISSVF--TKPDIAAPG 360
QY 396 AGVRGSGVGGCRALSGTSPVAVGAVTLVSTVQRELVPASMKQALIASARRLP 455
Db 361 VNITSTVRNGGYQAMSGSGSQASPTAGAVAVLISA---KPGASVDAIKNALFTSASNA 417

QY 456 VNMFEQGHGKLDL 468
| | | | |
DB 418 KNN-NVGFQGISI 429

RESULT 9

139780
subtilisin (EC 3.4.21.62) Sendai precursor - Bacillus sp.
C:Species: Bacillus sp.
C>Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 22-Jun-1999
C:Accession: I39780
R:Yamagata, Y.; Ishihara, K.; Ichishima, E.
Enzyme Microb. Technol. 17, 653-663, 1995
A:Title: Subtilisin Sendai from alkalophilic Bacillus sp.: molecular and enzymatic properties
A:Reference number: I39780; MUID:95329264; PMID:7605625
A:Accession: I39780
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-382 <RES>
A:Cross-references: GB:D29688; NID:9995963; PIDN:BAA06157.1; PID:9995964
C:Genetics:
A:Start codon: TTG
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:136-342/Domain: subtilisin homology <SBT>

Query Match 5.7%; Score 322; DB 2; Length 382;

Best Local Similarity 27.4%; Pred. No. 3.9e-15;
Matches 124; Conservative 59; Mismatches 172; Indels 98; Gaps 19;

QY 27 KKSKEAPCCSHLPLKVFSSVVEY-----YIVAFNGYFTAKARNSFTSSALK-SS 80
| | | | | :
DB 2 KLETKVVASA--LLLSISLTATSVSAEQKQVQLGIFENQLQV---TEFVSSDKQGS 56
QY 81 EVDNRIIPNPPSSDPSDFEVIQKEKAGLTLDHPNKRIPQPKVFSLKVAE 140
| | | | | :
DB 57 EMLFAEVNDESTEMELLVEFEDIPVVSVE---LSPEDVKLEK-DP-----SITVIE 105
QY 141 SD--PTVPCNETRWQKQSSRLRRASLSGSGFWHATGRHSSRRLLRAIPROVAQTLO 198
| | | | | :
DB 106 EDIEVTTNQVTPW-----GTHR-----VQ 125
QY 199 ADVLWQMGYTGANRVAVFDGLSEKHPKKNVKTWNT-NEETLDGLGHGTFAVGI 257
| | | | | :
DB 126 APTATRGYTGTVRVAVLDGIG-THPDL-NIRGVSVFVGFPSYQDNGRGTHTVATI 183
QY 258 ASMRE---COGFAPDAELHIFRFTNNQVSYTSWFLDAFNATIAILKIDVNLNLSIGPDM 314
| | | | | :
DB 184 AALNLSGVVGVAPNAELFAVILGANGSGSVSSIAQGLQTAONNIHVAFLSLGS----- 239
QY 315 DHPFDVKWELTAN-----NVIMVSAIGNDGPLYGLTNLPADQMDVIGVGGIDFEDNIA 369
| | | | | :
DB 240 --PVGSQTLAVNQATNAGVLVVAATGNNGS--GTVSYPARYANALAVGATDQNNRAS 295
QY 370 FSSRGMTTWELPGYGRMKPDIVTYGAGVRSKVGCGCRALSGTSVASPVVAGAVTLLVS 429
| | | | | :
DB 296 FSQGTGL-----NIVAPGVIGTQSTYGNRYASLSGTSNATPHVGAVALV-- 341
QY 430 TVQKRELVPASMKQALIASARRLPGVNMFEQ 462
| | | | | :
DB 342 -KQKNPSWNTQIRQLATSTATSLGNSNQPGSG 373

RESULT 10

SUBSN
subtilisin (EC 3.4.21.62) BPN' precursor - Bacillus amyloliquefaciens
N:Alternate names: subtilisin Novo
C:Species: Bacillus amyloliquefaciens
C>Date: 24-Apr-1984 #sequence_revision 28-Aug-1985 #text_change 21-Jul-2000
C:Accession: B25415; A93495; T44584; A92033; A00970
R:Vasantha, N.; Thompson, L.D.; Rhodes, C.; Banner, C.; Nagle, J.; Filpula, D.
J. Bacteriol. 159, 811-819, 1984

A:Title: Genes for alkaline protease and neutral protease from Bacillus amyloliquefac
A:Reference number: A25415; MUID:85006739; PMID:6090391
A:Accession: B25415
A:Molecule type: DNA
A:Residues: 1-382 <VAS>
A:Cross-references: GB:K02496; NID:9142525; PIDN:AAB05345.1; PID:9142526
A:Experimental source: ATCC 23844
R:Wells, J.A.; Ferrari, E.; Henner, D.J.; Estell, D.A.; Chen, E.Y.
Nucleic Acids Res. 11, 7911-7925, 1983
A:Title: Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin
A:Reference number: A93495; MUID:84069812; PMID:6316278
A:Accession: A93495
A:Molecule type: DNA
A:Residues: 1-382 <WE2>
A:Accession: T44584
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: W, 8-382 <WE2>
A:Cross-references: EMBL:X00165; NID:939337; PIDN:CAA24990.1; PID:9773560
R:Markland, F.S.; Smith, E.H.
J. Biol. Chem. 242, 5198-5211, 1967
A:Title: Subtilisin BPN'. VII. Isolation of cyanogen bromide peptides and the complete
A:Reference number: A92033; MUID:6808682; PMID:6065094
A:Accession: A92033
A:Molecule type: protein
A:Residues: 108-162, 'PN', 165-167, 'D', 169-194, 'SA', 197-204, 'DA', 207-264, 'ST', 267-357, '
R:Kraut, J.
In The Enzymes, 3rd ed., vol. 3, Boyer, P.D., ed., pp.547-560, Academic Press, New York
A:Title: Subtilisin: X-ray structure.
A:Reference number: A94443
A:Contents: annotation; X-ray crystallography, 2.5 angstroms; active site
C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and m
not necessary for normal sporulation.
C:Genetics:
A:Start codon: GTG
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-107/Domain: activation peptide #status predicted <APT>
F:108-382/Product: subtilisin BPN' #status experimental <EMPT>
F:130-342/Domain: subtilisin homology <SBT>
F:139,171,328/Active site: Asp, His, Ser #status experimental

Query Match 5.7%; Score 319.5; DB 1; Length 382;

Best Local Similarity 31.7%; Pred. No. 5.9e-15;
Matches 98; Conservative 50; Mismatches 118; Indels 43; Gaps 13;

QY 180 HSSRRLLRAIPROVAQTLOADVLMQGTGANRVAVFDGLSEKHPHFK-----NVKE 233
| | | | | :
DB 102 HVHAYAQSVPIGVSQ--IKAPALHSQGTGSNWKAVVIDSGDSSHPDLKVAGGASWPS 160
QY 234 RFTNWTNERTLDDGLGHGTFAVGVASMR---COGFAPDAELHIFRFTNNQVSYTSWFL 290
| | | | | :
DB 161 ETN-----PFQDNNSHGTHVAGTVAALNNSIGVLGVAPSASLYAVKVLGADGSGQYSMII 215
QY 291 DAFNYAILKIDVNLNLSIGGPDMD--HPFVDKRWELTANNVIMVSAIGNDGP--LYGTL 346
| | | | | :
DB 216 NGIEWAANNNDVNNINSLGSPGSAALKAAYDKA---VASGVVVAAGNESTSGSSSTV 272
QY 347 NNPAQMDVIGVGGIDFEDNIAIRFSRGMTTWELPGYGRMKPDIVTYGAGVRS--GVK 404
| | | | | :
DB 273 GPGKYPSTAVGAVDSSNQSRASFSGVGP---EL-----DVMAPGVSIQSTLPGNK 320
QY 405 GGCRAISGTSVASPVVAGAVTLLVSTVOKRELVPASMKQALIASARRLPGVNMFEQGHG 464
| | | | | :
DB 321 VG--AYNGTSMASPHVAGAAALILS---KHPNWTNTQVRSLSLENTTKLG--DSFYFGK 373
QY 465 KIDLILRAYQ 473
DB 374 LINVQAAQ 382

RESULT 11


```

Qy 453 LFGVNMFGQHGKLDL 468
Db 297 -----AQSGLLKL 305

RESULT 14
C41335
microbial serine proteinase (EC 3.4.21.-), intracellular - Bacillus polymyxa
C:Species: Bacillus polymyxa
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: C41335
R:Takekawa, S.; Uozumi, N.; Tsukagoshi, N.; Uedaka, S.
J: Bacteriol. 173, 6820-6825, 1991
A:Title: Proteases involved in generation of beta- and alpha-amylases from a large amylo
A:Reference number: A41335; MUID:92041565; PMID:1834632
C:Accession: C41335
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1336 <TAK>
A:Cross-references: GB:D00862; NID:g216285; PIDN:BA00735.1; PID:g216286
C:Superfamily: Bacillus intracellular serine proteinase; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:40-236/Domain: subtilisin homology <SBT>
F:49,86,244/Active site: Asp, His, Ser #status predicted

Query Match 5.3%; Score 299; DB 1; Length 326;
Best Local Similarity 30.3%; Pred. No. 1.3e-13;
Matches 94; Conservative 54; Mismatches 126; Indels 36; Gaps 12;

Qy 189 IPQVQATLQADVLQMGYTGANVRVAVFDTGLSEKHPHF-K-NYKRTNWTNERTID--- 244
Db 22 IPRGV-EMIQAPAVNQ-TRGRGVKVAVLDTGCDADHPDKARIIGRNFTDDGDEPEI 79

Qy 245 --DLGHTGVAGVIASMR---CQGEAPDAELHIFRVTNNQSVYTSWFLDAFNAILK 299
Db 80 FRDYNHGHTHVACTIAATENENGVVGVAPEADLLIKVLNKGSGQYDWIIQGIYATQ 139

Qy 300 KIDVLNLSIGDPMDHPV-DKWELTANNVMSAIGND---PLYGTLNAPADQMDV 355
Db 140 KVDIISNLSGPE--DVPDLHAKKAVASQILVWCAAGNEDGDDTDELGYPCYNEV 197

Qy 356 IQVGIDFEDNIARFSSRGMTWELPGYGRMKPDIVYAGVGRSGVKGCCRALSGTSV 415
Db 198 ISVGAINFDRHASEFN-----SNNEVDLVAPGEDILSTVPGKYATVSGTSM 245

Qy 416 ASPVYAGAVTL---LVSVQKRELVPNSMKAQALASARLP-GVNMFEQHGKLDLLRA 471
Db 246 APHPAGALALIKQLANASFENDUTEPELYAQLT---RTIPLGNSPKMGNGLLYLTA 302

Qy 472 YQLNSYKPK 481
Db 303 EELSRIFDTQ 312

RESULT 15
JC1085
subtilisin (EC 3.4.21.62) precursor - Bacillus licheniformis
N:Alternate names: alkaline proteinase
C:Species: Bacillus licheniformis
C>Date: 09-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 31-Mar-1997
C:Accession: JC1085
R:Lei, H.; Hong, Y.; Zhang, Y.Y.; Shen, T.J.
Chinese Biochem. J. 9, 441-447, 1993
A:Title: PCR amplifying cloning and sequencing of the coding sequence of the alkaline p
A:Reference number: JC1085
C:Accession: JC1085
A:Molecule type: DNA
A:Residues: 1-275 <LEI>
A:Note: The translation of the start codon ATG is not given in this paper
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:24-235/Domain: subtilisin homology <SBT>
F:33,64,221/Active site: Asp, His, Ser #status predicted

```

```

Query Match 5.3%; Score 295; DB 2; Length 275;
Best Local Similarity 31.5%; Pred. No. 2e-13;
Matches 93; Conservative 46; Mismatches 122; Indels 34; Gaps 9;

Qy 193 VAOT-----LOADVLQMGYTGANVRVAVFDTGLSEKHPHFKNKERTNWTNERTLDD 245
Db 1 MAOTVPYGIPLIKRADKVOAGGKGANVRVAVLDGTGQASHPDNLNVVVGASFAGEAYNTD 60

Qy 246 GLGHCTFFVAGVIASMR---GAPDAELHIFRVTNNQSVYTSWFLDAFNAILKKID 302
Db 61 GNGHCTHVAGTVAALDNTTGLGVAPSVSLYAVKLVNSGSGSYGIVSGIEWATTNGMD 120

Qy 303 VLNLSIGDPFMD--HPFDVKWELTANNVMSAIGNDPL--YGTLNAPADQMDVIGV 358
Db 121 VINLSIGSGSTAMKQAVNDAY---ARGVVVVAAGNSGSGNTNTIYPAKVDVJAV 177

Qy 359 GQIDFEDNIARFSSRGMTWELPGYGRMKPDIVYAGVGRSGVKGCCRALSGTSVASP 418
Db 178 GAVDSNSRASFSVSGA---EL-----EVNAPAGVYSTPTNYATLNGTSMASE 225

Qy 419 VVAGAVTLVSTVQKRELVPNSMKAQALASARLP-GVNMFEQHGKLDLLRAYQ 473
Db 226 HVAGAAALILS---KHPNLSASQVRNRLSSTATYLG--SSFYVGKGLINVEAAQA 275

RESULT 16
JC5460
intracellular alkaline serine proteinase (EC 3.4.-.-) - Thermoactinomyces sp.
C:Species: Thermoactinomyces sp.
C>Date: 17-Jun-1997 #sequence_revision 17-Sep-1997 #text_change 20-Jun-2000
C:Accession: JC5460; PC4328
R:Tsuchiya, K.; Ikeda, I.; Tsuchiya, T.; Kimura, T.
A:Title: Cloning and expression of an intracellular alkaline protease gene from alkal
A:Reference number: JC5460; MUID:97212029; PMID:9058969
C:Accession: JC5460
A:Molecule type: DNA
A:Residues: 1-321 <TSU1>
A:Cross-references: DBRI:DB5757; NID:g1536872; PIDN:BA13418.1; PID:g1536873
A:Experimental source: strain HS682
A:Accession: PC4328
A:Molecule type: protein
A:Residues: 16-45 <TSU2>
C:Superfamily: Bacillus intracellular serine proteinase; subtilisin homology
C:Keywords: hydrolase
F:16-321/Product: intracellular alkaline serine proteinase #status predicted <MAT>
F:40-263/Domain: subtilisin homology <SBT>
F:32,86,220/Active site: Asp, His, Ser #status predicted

Query Match 5.2%; Score 294; DB 2; Length 321;
Best Local Similarity 27.7%; Pred. No. 3e-13;
Matches 91; Conservative 54; Mismatches 111; Indels 72; Gaps 12;

Qy 186 LRAIPROVA-----OTLQADVLQMGYTGANVRVAVFDTGLSEKHPHFKNVK 232
Db 4 VRLIPEVTSIQDDTKKIPPTGEMTEAPDLWQOQYKGGIVVAVLDTGCDVEH---YELR 60

Qy 233 ERTNWTNERTLDDG-----LGHGTFVAGVIASMR---GAPDAELHIFRVTNN 281
Db 61 DELIKHNVTSDGDNDPELVSDQNGHGHVCGTIAATENDRAIGVAPCQLLVKVLNSR 120

Qy 282 QVSVTSWFLDAFNAIL-----KIDVLNLSIGDPFMDHPFDVKWELTANNVMSA 335
Db 121 GFGTEWVVGIRHAIWNGPGNGKAVQVLSNLSGKRE--NDPRLHDAIKENAVASGLVCA 179

Qy 336 IONGPLYGLTNLPADQM-----DVIGVGIDFEDNIARFSSRGMTWELPGYGRMK 388
Db 180 AGND-----GQGNSETDEFAYPGVPEVQVGSVLSGTSIRFSN-----SNCK 223

Qy 389 PDIVYAGVGRSGVKGCCRALSGTSVASPVVAGAVTLVSTVQK---RELVPNSMKAQ 445
Db 224 IDLVAPGEKILSTPGDKFATLTGTSMATPHVTGAALLTEKFEFERKITEPELPAQL 283

```


[illegible]

RESULT 20

S27501
alkaline proteinase (EC 3.4.21.-), intracellular - *Bacillus* sp. (strain 221)
C.Species: *Bacillus* sp.
C.Date: 10-Sep-1999
C.Sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C.Accession: S27501
R.Kato, C.; Nakano, Y.; Yamanoto, M.; Horikoshi, K.
submitted to the EMBL Data Library, March 1992
A.Description: Cloning and nucleotide sequence of the intracellular alkaline protease gene
A.Reference number: S27501
A.Accession: S27501
A.Molecule type: DNA
A.Residues: 1-321 <KAT>
A.Cross-references: EMBL: D10730; NID: g216233; PIDN: BAA01573.1; PID: g216234
C.Superfamily: *Bacillus* intracellular serine proteinase; subtilisin homology
C.Keywords: hydrolase; *Bacillus* intracellular serine proteinase; subtilisin homology
F.40-264/domain: subtilisin homology <SPT>
F.49.86.250/Active site: Asp, His, Ser status predicted

Query Match	5.1%;	Score 284.5;	DB 1;	Length 321;
Best Local Similarity	29.4%;	Prod. No. 1.4e-12;		
Matches	95;	Conservative 49;	Mismatches 122;	Indels 57; Gaps 14;
184	RLLRRAIP-ROV-----AQTLDQVLWQMGYTGANVRVAVFDTGLSEKHPEKFN	230		
Db	2	RKFRLLPQKQDKVSALSSEPMVGVEIAPAWKASAKAGQIIGVIDTGCQVDHP- --D	58	
QY	231	VKER-----TWNW-----ERTLDDGLGHGTFFVAGVIASMR-----CQGFAPDAELHIFRVF	278	
Db	59	LAERLIGGVNLTTDYGGVETFNDDNGHGTTHVAGTAAATGSGVGVYAPKADLFITKAL	118	
QY	279	TNNQVSYTSWFLDAFNAYI-----LKKIDVLNLSIGSGPDMDFPVDKWWELTANNVIM	332	
Db	119	SDGSSGEMGMIKAIKARYADVWRGPGEQEQRITTSGLSGPTDSEE-LHDAVKYAVSNNSV	177	
QY	333	VSAIGNDPLYTLNAN--PADQMDYIVGVGGIDFEDNIARFSRSGMTTWELPQGYGRMKP	389	
Db	178	VCAAGNEGDRQETEFAYPAATNEVIAVGAVDLRLSDFPN---TWEEI-----	225	
QY	390	DIIVTYGAGVGSVAGGCRALSGTSSVAPVAGATVLLTVSTVQ---KRELVNPSMKQAL	446	
Db	226	DIVAPGVGIKSTYSDSGVYAEISGTSMAAPHVAGALALIINLAKDAFKRTISETIECAQIV	285	
QY	447	IASARRLP-GVNNFPGHGKLDL	468	
Db	286	---RRATPIGTAQDKNGNFUTL	305	

RESULT 21

A09778 A precursor - Bacillus alcalophilus (stra
N; Alternate names: subtilisin homolog, high-alkaline
C; Species: Bacillus alcalophilus
C; Date: 12-Mar-1994 #sequence_revision 24-Feb-1995 #text_change 20-Jun-2000
R; Accession: A49778; JGI1244
R; Van der Jaan, J.C.; Gerritse, G.; Mulleners, L.J.S.M.; van der Hoek, R.A.C.; Quax,
Appl. Environ. Microbiol. 57, 901-909, 1991
A; Title: Cloning, characterization, and multiple chromosomal integration of a Bacillu
A; Reference number: A49778; MUID: I1282483; PMID:2059048
A; Cross-references:
A; Molecule type: DNA
A; Residues: 1-380 <V>
A; Cross-references: GB:M65086; NID:g142456; PIDN:AAA22212.1; PID:g142457
A; Experimental source: strain PB92, ATCC 31408
R; Nakami, H.; Kobayashi, T.; Kobayashi, M.; Yamamoto, M.; Nakamura, S.; Aono, R.; Hor
Biosci. Biotechnol. Biochem. 56, 1455-1460, 1992
A; Title: Molecular cloning, nucleotide sequence, and expression of the structural gen
A; Reference number: JCI244; MUID:93043753; PMID:1368952
A; Accession: JCI244
A; Molecule type: DNA
A; Residues: 1-195, 5', 197-380 <T>
A; Cross-references: GB:D13157; NID:g216231; PIDN:BAA02442.1; PID:g216232
A; Experimental source: Bacillus alcalophilus Vedder, ATCC 21522 (Bacillus sp. 221)
C; Superfamily: subtilisin; subtilisin homology
F; 1-27/pdomain: signal sequence #status predicted <S>
F; 28-111/pdomain: activation peptide #status predicted <PRO>
F; 112-380/product: alkaline serine proteinase #status predicted <MAT>
F; 134-340/pdomain: subtilisin homology <SBT>
F; 143,173,326/Active site: Asp, His, Ser #status predicted

```

Query Match          5.08; Score 283; DB 2; Length 380;
Best Local Similarity 25.9%; Pred. No. 2.4e-12;
Matches 113; Conservative 65; Mismatches 165; Indels 94; Gaps 17;

41  LTLKVFSSV-----VEYEVAFNGVFTAKARNFSISALKSSEVDNWRILPRNNPS 94
   | : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
14  LLTSVAFSSASTASAEBAEKYLTGFN---EQEAVSEFVEQVEANDV---ALLSEE--- 64
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
95  SDVPSDFEVIQKKQKAGLLTLEDPH-----NIKRVTPQRKVFRLSKYASDTPVPCNTR 151
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
65  -----EVEITLHFETIPVLSVELSPEDVDALDLP-----AISTEEDAEV-----TT 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
152  WSKQ---WOSSRPLRRASLSGSGFWHATGRHSSRRLLRAIPROVATLQADVLMQVMTG 209
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
111  MAOSVPMWGISR-----VOQPAAHNRGLTG 134
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
210  ANYRVAVFDTLGSEKHFPKVKVKNRTNWT-NERTLDDGLGHGTFAVGIASMRE---CQG 265
   | : | | | | : | : | : | : | : | : | | | | | | | | | | : | : |
135  SGYKVAVLDTGIS-THPDL-NIRGGASFVPGEPSTQDNGCHGTHTAGTIALNNISGVLG 192
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
266  FAPDAELHFRVFTNNQVSTSWFLDAFNAILKKIDVLNLSIGGPDFMDHPFDVKWEL 325
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
193  VAPNAELAVKVLGASGSGSVSSIAQLEWAGNMGHMANLSLGSF-PSATLEQAVNSA 251
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
326  TANNVIMVSAIGNDPGLYGLNNPADQMDVIGVGIGDFEDONIAFSSRGMTTWELPGYG 385
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
252  TSRGVLVVAASGNSGA---GGSISYPARYANAAGVATQDNNRASFSQYGAGL----- 301
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
386  RMKPDILTVYGVAGVSGWGKCGCRALSGTGSVAPVAVGATLLVSTYQKRELVPASMKQA 445
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
302  ----DTPAVGNVOSTYPGSTYASLNGTSMATPHVGAALV---KQKNPSWSNVQIRNH 354
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
446  LIASARRLPGVNMFEOG 462
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
355  LKNTATSLGSLNLYSGG 371
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 22
T28159

pyrolysin (EC 3.4.-.-) - Pyrococcus furiosus
C;Species: Pyrococcus furiosus
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 24-Oct-2000
C;Accession: T28159
R;Voorhorst, W.G.B.; Eggen, R.I.L.; Geerling, A.C.M.; Platteeuw, C.; Slezzen, R.J.; Vos, J. Biol. Chem. 271, 20426-20431, 1996
A;Title: Isolation and characterization of the hyperthermostable serine protease, pyrolysin
A;Reference number: Z20481; MUID:96355370; PMID:8702780
A;Accession: T28159
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1398 <VOO>
A;Cross-references: EMBL:U55835; NID:g1556462; PID:g1556463; PIDN:AAB09761.1
A;Experimental source: DSM3638
C;Genetics:
A;Gene: pls
C;Keywords: hydrolase; serine proteinase

Query Match 5.0%; Score 281.5; DB 2; Length 1398;
Best Local Similarity 22.9%; Pred. No. 2.5e-11;
Matches 164; Conservative 84; Mismatches 222; Indels 245; Gaps 34;
QY 245 DGLGHGTFVAGVTA-----SMRECOGFAPDAE 271
DB 361 DGHGHGTHVAGVAGYDSDNDAWDLMSYGEWEVSRLYGWDYTNVTDVQGVAPQA 420
QY 272 LHFIRVETNNQVSYTFLDAFNAILKIDVLNLSIGGDFMDHPVD-----KWVE 324
DB 421 IMAIRVLRSDGRSMWDIIIEGMYAATHGADVLSMLGG---NAPYLDGTDPSVADE 476
QY 325 LTAN-NVIMVSAIGNDGLYGTLLNPNADQMDVIGV-----GIDFED----- 365
DB 477 LTEKYGVWFVIAAGNEGPGINVGSPGATKALTGAAAVPINVGVVVSQALGPDYIGF 536
QY 366 -----NIAFSSRGMTTWELPGCYGRMKPDIVTYGAGVRSQVYK---GGCRALSGTS 414
DB 537 YFPATNVRIAFSSRG-----PRIDGEIKPNVAPGYIYSSLPWMVGGADFMSTGS 590
QY 415 VASPVVAGATLLVSTVQKREL-VNPSMKOALIASARLPG-----VNMEEQHGKL 466
DB 591 MATPHSVGVVALLISGPKPGIYINPDIIKKVLESGATWLEGDPYTCQKYKTELDOGHGLV 650
QY 467 DLRAVQILNSYKQASLSPSYIDLTECPYMWYCSQPIYGMPTVVNVVTLNMGVGTG 526
DB 651 NVTKSWEILKA-----INGTLPITVDHWADKS--YSDFAEYLGVDVIRGLYARN 697
QY 527 RIVDKPDWQYLPQNGDNIENAVSYSSVLWPSGLAISVTKKAASW-EGTAQGHVMI 585
DB 698 SIPDIVEW--HIKYVD-----TEYRTEIYATE---PWIKPFVSGSVIL 737
QY 586 TVASPAETESKGAETSTVKLPK-----VKII---PTPPSRKRVLDQVHNLRYP 634
DB 738 -----ENTEFVLRYKYDVEGLEPGLYVGRILIDPTTP-----VIEDELLNTIVI 783
QY 635 PGTFPRDLNRMKNLDLW-----NGDHIHINFRDMYQHLRSMGYFVEVLGAPFTCFDASQY- 690
DB 784 PERKTPEN-----NVLTLWYDINGPEMVTH-----HFTVPEGVDVLYAMTVDYGLYR 833
QY 691 --GTLMLVDSEEEYFPEETAIAKLRRVDNGLSLVIFSDWNTYVMRKVKFDENTROWMWP 748
DB 834 PGMFV-----FPYQDYLPAAVSNMP-----GNW----- 859
QY 749 DTGGANIPALNELLSVNMGMFSDG-LYEGEFTLANHDMYVYASGCSIAKFPEDGVILT--- 804
DB 860 -----EL--VWT-GFNAPLYESGFLVRIH-----GVEITPSV 889
QY 805 -----QTFKQGLEVLKQETAVVENVPI-----LGLYQIPAE--GGGRIVLYG 845
DB 890 WYINRYLDNTFEESIEFNITNIYAPINATFLIPGLGTYNASVESVDGDEFFIKG 944

RESULT 23
JC4802

alkaline proteinase (EC 3.4.21.-) precursor - Thermoactinomyces sp. (strain E79)
C;Species: Thermoactinomyces sp.
A;Variety: strain E79
C;Date: 15-Aug-1996 #sequence_revision 15-Oct-1996 #text_change 09-Jun-2000
C;Accession: JC4802
R;Lee, J.K.; Kim, Y.O.; Kim, H.K.; Park, Y.S.; Oh, T.K.
Biosci. Biotechnol. Biochem. 60, 840-846, 1996
A;Title: Purification and characterization of a thermostable alkaline protease from T
A;Reference number: JC4802; MUID:96261070; PMID:8704314
A;Accession: JC4802
A;Molecule type: DNA
A;Residues: 1-384 <LEE>
A;Cross-references: GB:U31759; NID:g1389689; PIDN:AAB36499.1; PID:g1683629
A;Experimental source: strain E79
C;Comment: This protein is thermostable.
C;Function:
A;Description: extracellular alkaline serine proteinase [validated, MUID:96261070]
C;Superfamily: subtilisin; subtilisin homology
C;Keywords: hydrolase; serine proteinase
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-106/Domain: propeptide #status predicted <PRO>
F;107-383/Product: alkaline proteinase #status experimental <MAT>
F;134-344/Domain: subtilisin homology <SBT>
F;143-176/330/Active site: Asp, His, Ser #status predicted

Query Match 4.9%; Score 276.5; DB 2; Length 384;
Best Local Similarity 24.0%; Pred. No. 7.1e-12;
Matches 115; Conservative 75; Mismatches 171; Indels 119; Gaps 20;
QY 9 LLLVVLGCKKHLDRLEKSKFEKAPCGGSHLTLYKVEFSSTVVEY---EYVAFNGYFT 65
DB 11 LLLVLLA-----VPG-----TMFAASPASTDYDYPGELIVFKDGIS 48
QY 66 AKARNSFISS-ALKSSEVDNWRRIIPRNPSSDYSPSEFVQIKKQAGLLTLEDHNIK 124
DB 49 AQSTQSTHAQYGAKSIE-----KSKY-LGFEVVKFDGSEVMKIEKYKNNPNVE 95
QY 125 RVTQPRKFRSLKYAESDPTVPCNETRWSOKWQSRPLRSLASLSGSGFWHATGRHSRR 184
DB 96 YVEPNHYV--HIMTTPNDLT-----SRQWG----- 118
QY 185 LLRAIPROVAQTLOADVLMQYGTGANVRVAVFTGLSEKHPH--KNVKERTWNTWERT 242
DB 119 -----PQVQAPQAWDVTSSSTVIAIVDTGVQTNHPDQLQGIQVQDYDFVDSN 169
QY 243 LDGLGLHGTFFVAGVIASMR-----CQGFAPDAELHIFRVFTNNQVSYTSFELDAFNAIL 298
DB 170 PQDNGHGTCHAGIAAAVATNNGTGIACMAPNASIMPVRLNNSGSGTMAAVANGIAVAAQ 229
QY 299 KKIDVLNLSIGGPDFMD--HPFVDKVMELTANNVIMVSAIGNDGLYGTLLNPNADQMDVI 356
DB 230 NGADVLSLGGTSGSALQSAVQOAWN---SGAVVVAAGNSS--STPNYPAYISOAI 284
QY 357 GVGIDPEDNIARFSSRGMTTWELPGCYGRMKPDIVTYGAGVRSQVGGCRALSGTSVA 416
DB 285 AVATSDNSLSYFSNYG--SW-----VDVAAPGSGNIYTYLNSVYASLSGTSMA 332
QY 417 SPVVAGAVTLLVSTVQRELNVNPSMKOALIASARLPGVNMFFQGHKLDLLRAYQILN 476
DB 333 TPHVAGIALLAS--QGR---SNSQIRAAIENTADKISGTGYTFQ-HGRIN---AYKAVN 383

RESULT 24
G95392

probable protease [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSyna
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: G95392
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli
A;Reference number: A95262; MUID:21396509; PMID:11481432

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 2, 2003, 02:50:54 ; Search time 113 seconds
(without alignments)
2855.079 Million cell updates/sec

Title: US-09-830-837-6
Perfect score: 5617
Sequence: 1 MKLVNIVLLVLLCGKKH.....PRVKRPLMQQVHPKPTPSV 1052

Scoring table: BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DRV=xlh
-O=/cgn2_1/USPTO.spool/US09830837/runat_23052003_181925_8091/app_query.fasta.1.1223
-DB=Issued_Patents_NA -QFMT=fastp -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=25
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830837.ecgn_1.1.46 -runat_23052003_181925_8091 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	5617	100.0	4338	4	US-09-360-237-4	Sequence 4, Appl1
2	5487	97.7	4198	4	US-09-360-237-2	Sequence 2, Appl1
3	338.5	6.0	1500	1	US-08-322-965-1	Sequence 1, Appl1
4	331	5.9	1497	1	US-08-322-677A-6	Sequence 6, Appl1
5	331	5.9	1497	1	US-08-322-676-6	Sequence 6, Appl1
6	331	5.9	1497	3	US-08-898-218-6	Sequence 6, Appl1
7	331	5.9	1497	3	US-08-848-793-6	Sequence 6, Appl1
8	331	5.9	1497	4	US-09-445-270-1	Sequence 1, Appl1
9	331	5.9	1513	4	US-09-178-155-1	Sequence 1, Appl1
10	330	5.9	2809	3	US-09-000-016-1	Sequence 1, Appl1
11	330	5.9	2809	4	US-09-514-340-1	Sequence 1, Appl1
12	329	5.9	1496	6	RE34606-1	Patent No. RE34,606

13	328.5	5.8	1494	4	US-09-255-502-1	Sequence 1, Appl1
14	328.5	5.8	1497	6	5472855-1	Patent No. 5472855
15	328.5	5.8	1868	1	US-08-069-863-1	Sequence 1, Appl1
16	328.5	5.8	1868	1	US-08-309-069-1	Sequence 1, Appl1
17	328.5	5.8	1868	5	US-08-562-833-1	Sequence 1, Appl1
18	328.5	5.8	1868	5	US-08-562-833-1	Sequence 1, Appl1
19	328.5	5.8	1868	5	US-08-562-833-1	Sequence 1, Appl1
20	328.5	5.8	1868	5	US-08-562-833-1	Sequence 1, Appl1
21	328.5	5.8	8119	1	US-08-460-343B-1	Sequence 1, Appl1
22	328.5	5.8	8119	1	US-08-398-028B-1	Sequence 1, Appl1
23	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
24	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
25	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
26	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
27	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
28	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
29	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
30	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
31	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
32	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
33	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
34	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
35	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
36	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
37	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
38	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
39	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
40	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
41	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
42	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
43	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
44	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1
45	328.5	5.8	8119	2	US-08-504-265B-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-360-237-4
; Sequence 4, Application US/09360237
; Patent No. 6322962
; GENERAL INFORMATION:
; APPLICANT: BROWN, MICHAEL S.
; APPLICANT: CHENG, DONG
; APPLICANT: ESPENSHADE, PETER J.
; APPLICANT: GOLDSTEIN, JOSEPH L.
; APPLICANT: RAWSON, ROBERT B.
; APPLICANT: SAKAI, JUKO
; TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF
; FILE REFERENCE: UTXD:567
; CURRENT APPLICATION NUMBER: US/09/360,237
; CURRENT FILING DATE: 1999-07-23
; EARLIER APPLICATION NUMBER: 60/096,571
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Human
US-09-360-237-4

Alignment Scores:
Pred. No.: 0
Score: 5617.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 4
Length: 4338
Matches: 1052
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-830-837-6 (1-1052) x US-09-360-237-4 (1-4338)

Qy	1	MeLysLeuValAsnIleTrpLeuLeuLeuValValLeuLeuCysGlyLysLysHis	20		1577	ATTGACTTTGAAGATTAACATCGCCGCTTTCTTCAAGGGGAATGACTACCTCGGAGCTA	1636
Db	497	ATGAAGCTTGTCAACATCTGGCTGCTCTGCTCGTGGTTTGTCTCTGGGAAGAAACAT	556		381	ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly	400
Qy	21	LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis	40		1637	CCAGAGGCTACGGTCGCATGAACCTGCATTTGCACATATGGTGTCTGGCGTGGGGGT	1696
Db	557	CTGGCGCAGACACTGGAAAGAAATCTTTTGAAGAGCCCATGCCCTGGCTGTCCCCAC	616		401	SerGlyValLysGlyGlyCysArgAlaLeuSerSerGlyThrSerValAlaSerProValVal	420
Qy	41	LeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrIleValAlaPhe	60		1697	TCTGGCTGAAAGGGGGTCCGGGCCCTCTCAGGACACAGTGTGTCTCTCCAGTGGTT	1756
Db	617	CTGACTTTGAAGGTGGAAATCTCATCAACAGTTGTGGAATATGAATATATGTGGCTTTC	676		421	AlaGlyValAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuLeuValAsnProAla	440
Qy	61	AsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSer	80		1757	GCAGTGTGTCACTTGTAGTGAGCACAGTCCAGAAAGCGTGAGCTGTGTAATCCCACC	1816
Db	677	AATGGATATTTACAGCAAAGCTAGAAATTCATTTATTTCAGGTGCCCTGAAGAGCAGT	736		441	SerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAlaAsnMetPheGlu	460
Qy	81	GluValAspAsnTrpArgIleIleProArgAsnAsnProSerSerAspTyrProSerAsp	100		1817	AGTATGAAGCAGGCCCTGTATCGCTCAGCCCGAGGCTCCCGGGGTCAACATGTTTCAG	1876
Db	737	GAAGTAGACAATTGGAGAAATTATACCTCGAAACAATCCATCCAGTCAGTACTACCTAGTCAT	796		461	GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro	480
Qy	101	PheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis	120		1877	CAAGCCACGGCAAGCTCGATCTGCTCAGAGCTTATCAGATCCTCAACAGCTACAAGCCA	1936
Db	797	TTTGAAGTGAATTCAGATAAAGAAACACAGAAAGCGGGCTGCTAAACATCAAGATCAT	856		481	GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr	500
Qy	121	ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu	140		1937	CAGGCAAGTTTGGCCCCACGCTACATAGATCTGACTGAGTGTCCCTACATGTGGCCCTAC	1996
Db	857	CMAACATCAACCGGTCAGCCCCCAACGAAAGTCTTCTCTCCCTCAAGTATGCTGAA	916		501	CysSerGlnProIleTyrTyrGlyMetProThrValValAlaAsnValThrIleLeuAsn	520
Qy	141	SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg	160		1997	TGCTCCCGACCATCTACTATGGAGGAATGCCGACGTTGTATATGTCACATCTCAAC	2056
Db	917	TCATGCCCAACAGTACCTTGCAATGAAACCCGGTGGAGCGAGAGTGGCAATCATCAGT	976		521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln	540
Qy	161	ProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis	180		2057	GGCATGGAGTCAAGGAAGAAATGTAGATAAGCCCTGACTGGACGCCCTATTGTCACAG	2116
Db	977	CCCTCGAAGACGCCCTCTCCCTGGGCTCTGGCTCTGGCATGCTACGGGAAGGCAT	1036		541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGly	560
Qy	181	SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp	200		2117	AACGGAGACAACATTTGAAGTGGCTTCTCTACTCCTCGGCTTATGGCCCTTGGTCGGGC	2176
Db	1037	TCGAGCAGACGGCTGTGAGAGCCATCCCGCGCCAGGTGGCCAGACACTGCGAGGCAGAT	1096		561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln	580
Qy	201	ValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGly	220		2177	TACTGGCCATCTCCATTTCTGTGACCAAGAAAGCGGCTCTCTGGGAAGGCATTTCTCTCAG	2236
Db	1097	GTCTCTGGCAGATGGGATATACAGGTGCTAAATGTAAAGATGTGCTTTTTCACACTGGG	1156		581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600
Qy	221	LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu	240		2237	GGCCATGTATGATCCTCTGTGGCTTCCCGCAGCAGACAGACAGTCAAAAAATGGTCAGAA	2296
Db	1157	CTGAGCGAAGACATCCCACTCAAAATGTGAAGAGAGAACAACCTGGACCAACGAG	1216		601	GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProArgSer	620
Qy	241	ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet	260		2297	CAGACTCAACAGTAAAGCTCCCAATTAAGGTGAAGATAATTCTACTCCCCCGCGAAGC	2356
Db	1217	CGAACCTGGACGATGGGTGGGCCCATCGGCACATTCGTGGCAGGTGTGATAGCCAGCATG	1276		621	LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg	640
Qy	261	ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn	280		2357	AAGAGAGTTCTCTGGATCAGTACCACAACCTCCGCTATCCACTGGCTATTTCCTCCAGG	2416
Db	1277	AGGGAGTGCAGAGATTGCTCCAGATCGAACAATTCACATTTTCAGGGTCTTTTACCAAT	1336		641	AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660
Qy	281	AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLys	300		2417	GATAATTTAAGCATCAAGATGACCCCTTTAGACTGGAATGCTGATCACAATCCACACCAAT	2476
Db	1337	AATCAGGATCTTACACTCTTGGTTTTTGGAGCCCTTCAACTATGCCATTTTAAAGAAG	1396		661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
Qy	301	IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisProPheValAsp	320		2477	TTCAAGGATATGTACAGCATCTGAGNACATGGGCTACTTTGTAGAGTCTCTCGGGGCC	2536
Db	1397	ATCGACGTGTTAAACCTCAGCATCGCGGCCGAGCTCATGGAATCATCGTTTGTGAC	1456		681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
Qy	321	LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly	340		2537	CCCTTCACGTGTTTGTATGGCCAGTCAAGTATGGCACTTGTGTAGTGGACAGTGAAGG	2596
Db	1457	AAGGTGGGAATTTACAGCTAACCAATGTAAATCATGGTTTCTGTATTTGGCAATACCGA	1516		701	GluTyrPheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu	720
Qy	341	ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly	360		2597	GAGTACTCTCCCTGAAAGAGATCGCAAGCTCCGAGGAGGACGTGGACAACCGCTCTCGCTC	2656
Db	1517	CTCTTTATGGCACTCTGAATTAACCTGCTGATCAAAATGGATGTGATTGGATAGCGGC	1576		721	ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
Qy	361	IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThrTrpGluLeu	380				

```
Db 2657 GTGATCTTCACTGACGTGGTACAACTCTCTGTTATGAGAAAAGTGAAGTTTATGATGAA 2716
QY 741 AsnThrArgGlnTrpMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu 760
Db 2717 AACACAGGAGTGGTGGATGCGGATACCGGAGAGCTAACATCCAGCTCTGAATGAG 2776
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrgluGlyPheThrLeu 780
Db 2777 CTGCTCTCTGTGTGAACATGGGTTCAGCGATGGCTGTATGAAGGGAGTTCACCCCTG 2836
QY 781 AlaAsnHisAspMetTyrrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
Db 2837 GCCAACCATGACATGATTATGCGTCAGGTCAGCATCGCAAGTTTCCAGAAGATGGC 2896
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
Db 2897 GTGCTGATACACAGACTTTCAGGACCAAGGATGGAGTTTAAAGCAGGAAACAGCA 2956
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrglnIleProAlaGluGlyGlyArg 840
Db 2957 GTTGTGTGAAAACGTCCCATTTTGGACATTTATCATCTCCAGCTCAGGTCGAGGCCGG 3016
QY 841 IleValLeuTyrglyAspSerAsnCysLeuAspSerHisArgGlnLysAspCysPhe 860
Db 3017 ATTGTACTGTATGGGACTCCCAATTCGTTGGATGACAGTCACCGACAGAAGGACTGCTT 3076
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrrThrSerTyrglyValThrProProSerLeuSer 880
Db 3077 TGGCTTCTGGATGGCCCTCTCCAGTACACATCTATGGGGTGACACCGCTAGCCCTCAGT 3136
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
Db 3137 CACTCTGGGAACCGCCAGCCGCTCCAGTGGAGCAGGCTCAGTCACTCCAGAGAGATG 3196
QY 901 GluGlyAsnHisLeuHisArgTyrrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
Db 3197 GAAGGAAACCATCTTCCATCGGTACTCCAAAGGTTCTGGAGGCCCATTTGGGAGACCCAAA 3256
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
Db 3257 CCTGGCCCTTACCAGCCTGTCCACGCTGTCTTGGGCCAGCCACAGCCTTTAAACGAG 3316
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLysVal 960
Db 3317 ACGGCCCCAGTAACCTTTGGAAACATCAGAAGCTACTCCCATTCACCTGGACAAAGTG 3376
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3377 GTGTTACCCCACTTTCGATCGAATCGCCCTCAAGTGAGGCCCTTGTCCCTGGAGAGAGC 3436
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrrAsnGlnGluValGlyGln 1000
Db 3437 GCGCCTGGGACATCTCTGGAGGATCATGCTGGCCGCTACACACAGAGGTGGCCAG 3496
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
Db 3497 ACCATTCTCTCTTTCCTCTCTGGAGCATGTGGTCTCGCTCTCTTGTGTGTAACA 3556
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3557 ATCAACAGGCCCAAGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3616
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3617 ATGCAGCAGGTTACCCGCCCAAGAGACCCCTCTGGTG 3652
```

RESULT 2

US-09-360-237-2

; Sequence 2, Application US/09360237

; Patent No. 6322962

; GENERAL INFORMATION:

; APPLICANT: BROWN, MICHAEL S.

; APPLICANT: CHENG, DONG

```
; APPLICANT: ESPENSHADE, PETER J.
; APPLICANT: GOLDSTEIN, JOSEPH L.
; APPLICANT: RAWSON, ROBERT B.
; APPLICANT: SAKAI, JURO
; TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF
; FILE REFERENCE: US/09-360, 237
; CURRENT APPLICATION NUMBER: 1999-07-23
; EARLIER APPLICATION NUMBER: 60/096, 571
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 4198
; TYPE: DNA
; ORGANISM: Cricetus griseus
; US-09-360-237-2
```

Alignment Scores:

Pred. No.:	0	Length:	4198
Score:	5487.00	Matches:	1021
Percent Similarity:	98.57%	Conservative:	16
Best Local Similarity:	97.05%	Mismatches:	15
Query Match:	97.69%	Indels:	0
DB:	4	Gaps:	0

US-09-830-837-6 (1-1052) x US-09-360-237-2 (1-4198)

QY	1	MetLysLeuValAsnIleTrpLeuLeuValValLeuLeuCysGlyLysLysHis	20
Db	387	ATGAAGCTCATCAACATCTGGCTTCTTCGTGGTGGTTTGTCTGTGGAAAGAGCAT	446
QY	21	LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis	40
Db	447	CTGGGTGACAGGCTGGGGAGAAAGCGTTTGAAGAGGCATCATCGCCTAGCTTCCAC	506
QY	41	LeuThrLeuLysValGluPheSerSerThrValValGluTyrrGluTyrrIleValAlaPhe	60
Db	507	CTGACTTTCGAAGTGGAAATCTCTCACTGGTGGATATGAATATATTGTGGCTTC	566
QY	61	AsnGlyTyrrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSer	80
Db	567	AACGGATACTTCACAGCAAGCTAGAACTCATTTATTTCAAAGTGTCTGAAAGCAGT	626
QY	81	GluValAspAsnTrpArgIleIleProArgAsnAsnProSerSerAspTyrrProSerAsp	100
Db	627	GAAGTAGACAACCTGGAGAATTATACCTCGGAACAACCCATCCAGTACCTAGTATGAT	686
QY	101	PheGluValIleGlnLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis	120
Db	687	TTTGAGTGTATTCAGATAAAGAGAGAGAGAGCGGGCTGCTCACACTTGAAGATCAT	746
QY	121	ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrrAlaGlu	140
Db	747	CCAAACATCAAGCGGGTGACACCTCAACGCAAGTCTTTCGTTCCTTGAAGTTTGTGAA	806
QY	141	SerAspProThrValProCysAsnGluThrArgTyrrSerGlnLysTrpGlnSerSerArg	160
Db	807	TCTGACCCCATTTGCGCATGTATGAACCTCGGTGGAGCGAGAGTGGCAGTCATCACA	866
QY	161	ProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis	180
Db	867	CCCTGAGAGAGAGCCAGTCTCTCCCTGGGGCTCTGGATTCGGCATTCACAGAGAACAT	926
QY	181	SerSerArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp	200
Db	927	TCAAGCCGCGATTGTGTGAGAGCCATTCCTCGACAGGTTGCCAGACATTCGAGGCAGAT	986
QY	201	ValLeuTrpGlnMetGlyTyrrThrGlyAlaAsnValArgValAlaValPheAspThrGly	220
Db	987	GTCTGTGGCAGATGGGATACAGAGTGTAAATGTCAAGGTTCTGCTGTTTGTACTGGG	1046

QY	221	LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu	240
Db	1047	CTCAGTGAAGACATCCACATTTCAAGAAATGTGAAGAGAACCAACTGGACCAATGAG	1106
QY	241	ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet	260
Db	1107	CGAACCTCGGATGATGGGTGGCCATGGCACATTTGTGCGAGGTGTGATGTCACGACATG	1166
QY	261	ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn	280
Db	1167	AGGGAGTGCACAGGATTTGCCACAGATGCAGAGCTGCACATCTTCCGGGTCTTTACCAAC	1226
QY	281	AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLys	300
Db	1227	AATCAGGTGTCTTACACATCTTGGTTTGTGGAGCTTCAACTATGCCATCTCAAGAAAG	1286
QY	301	IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisProPheValAsp	320
Db	1287	ATTGATGTCTTAAACCTTATGACATCGCGGGCTGTACTTCATGGATCATCCCTTGTTCAC	1346
QY	321	LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly	340
Db	1347	AAGGTGTGGGAATTAACAGCTAACAAATGTAATCATGGTTTCTGTATCGGCAATGATGA	1406
QY	341	ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyGly	360
Db	1407	CTCTTTATGGCACTCTGAATTAACCCAGCTGATCAGATGGATGTGATGGAGTGGGTGGC	1466
QY	361	IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGluLeu	380
Db	1467	ATTGACTTTGAAGATAACATCCGCCCTTTCTTCCGAGGAATGACTACTTGGGAACATA	1526
QY	381	ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly	400
Db	1527	CCAGGAGGCTATGTGCGGTGAACCTGCAATTTGTCACTATGTGCGGAGTGGCGGGT	1586
QY	401	SerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal	420
Db	1587	TCCGGTGTGAAGGGGGCTCGCGGCACTCTCAGGAGACCATGTCTGCCCTTCCCGAGTGGTT	1646
QY	421	AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAla	440
Db	1647	GCTGGGGCTGTACCTTGTATTAGACACAGTCGCAAGCGGGAGCTAGTGAATCCTGCC	1706
QY	441	SerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAlaAsnMetPheGlu	460
Db	1707	AGTGTGAAGCAAGCCCTGATTCTCATCGCCGGAGGCATCTCTGGTGTTAACATGTTCCAG	1766
QY	461	GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro	480
Db	1767	CAAGGCCATGCAAGCTGATCTGCTCGAGCCTATCATAGATCTCTCAGCAGCTACAAACA	1826
QY	481	GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr	500
Db	1827	CAGGCGAGCTTGACTCTAGCTACATCGACCTCAGCTGATGTCCCTACATGTGGCTTAC	1886
QY	501	CysSerGlnProIleTyrTrpGlyCysMetProThrValValAsnValThrIleLeuAsn	520
Db	1887	TGTTCTCAGCCCATCTACTATGAGGAATGCCAACTTTGTAATGTACACATCTCAAT	1946
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln	540
Db	1947	GGCATGGGAGTCACAGAAATTTGTGGTAAGCTGAGTGGCGGGCTTATTATTCACAG	2006
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGly	560
Db	2007	AATGGAGACAACATTTGAATGGCTTCTCTACTCTCTCAGTGTATTGGCTTGGTCAGGC	2066
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln	580
Db	2067	TACCTGGCCATCTCCATTTCTGTCAACCAAGAGCAGCTCTCTGGGAAGGCATTCACAG	2126
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600

Db 3207 ACTGCGCCAGTATCTTTGGAAACATCAGAGCTGCTCTCCATTCACCTGGAGAAAGTA 3266
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3267 GTGTTACCCAACTTCATCGAATCGCCCTCAAGTAGAGACCTTTGTCCCTGGAGAAAGT 3326
QY 981 GYAlaTrpAspIleProGlyGlyIleMetProGlyArgTrpAsnGlnGluValGlyGln 1000
Db 3327 GGTGCTGGGACATCTCTGGAGGATCATGCTTGGCGCTGACCAACAGAGGTGGGCCAG 3386
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGln 1020
Db 3387 ACCATCCCTGCTCTTTCCTCTCCCTCGGAGCCATGCTGGCCCTCTTCTTGTGTACAG 3446
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgAlaGlyProArgValLysArgProGlnLeu 1040
Db 3447 ATCAGCAAGCCAAAGCGCGCGAAGGAGGAGGCCAGGCCAAGAGCGTCCACAGCTT 3506
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3507 ACACAGCAGACCCACCACCAAGGACCCCGTCAAGT 3542

RESULT 3
US-08-322-965-1
; Sequence 1, Application US/08322965
; Patent No. 5733473
; GENERAL INFORMATION:
; APPLICANT: Johnston, James P.
; APPLICANT: Lenoir, Pierre M.A.
; APPLICANT: Thoen, Christian A.J.K
; TITLE OF INVENTION: Liquid Detergent Composition Containing
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 5299 Spring Grove Avenue
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: USA
; ZIP: 45217
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,965
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 252
; ATTORNEY/AGENT INFORMATION:
; NAME: Allen, George W.
; REGISTRATION NUMBER: 26,143
; REFERENCE/DOCKET NUMBER: CM-352MC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 513/627-5946
; TELEFAX: 513/627-8118
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 455..1282
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 137..1282
US-08-322-965-1

Alignment Scores: 1.53e-23 Length: 1500
Pred. No.:

Score: 338.50 Matches: 133
Percent Similarity: 40.58% Conservative: 76
Best Local Similarity: 25.83% Mismatches: 180
Query Match: 6.03% Indels: 126
DB: 1 Gaps: 21
US-09-830-837-6 (1-1052) x US-08-322-965-1 (1-1500)
QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
Db 233 AGCAGTACA-----GAAAGAAATACATTTGTCGGATTT-----AAA 268
QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAlaAspAsnTrpArgIle 87
Db 269 CAGACATAGTCCATGAGTTCGCCAAGAAAAGATGTTAT----- 313
QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
Db 313 ----- 313
QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
Db 314 ----TCTGAAAAGGCGGAAAGTTCAAAAGCAATTTAAGTATGTTAACGCGCGCCAGCA 370
QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
Db 371 ACATTGGATCAAAAAGCTGTAAAGAAATTGAAA-----AAAGATCCGACGGTTCGA 421
QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSer 166
Db 421 ----- 421
QY 167 LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
Db 422 -----TATGTGGAAGAAGATCATATTCACATGAATATGCG 457
QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
Db 458 CAATCTGTCTTATGCGATTCTCAA---ATTAAAGCGCGGCTCTTCACTCTCAAGGC 514
QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
Db 515 TACACAGGCTCTAACGTAAAGTAGCTGTATGACAGCGGAATGACTCTTCTCATCCT 574
QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
Db 575 GACTTA---AACGTACAGCGCGGAGCAAGCTTCGTACTCTTCTGAAACAAACCCATACCAG 631
QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 632 GACGGCAGTCTCACGGTAGCGCATGTAGCGGTACGATGCGGCTCTTAATACTCAATC 691
QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
Db 692 GGTGTTCTGGCGGTAGCCCAAGCGCATCATATATGACAGTAAAGTGTGTGATTAACA 751
QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysIle 301
Db 752 GGAAGCGGCAATATAGCTGGATTATTACGGCATTGAGTGGCCATTCCCAACAATATG 811
QY 302 AspValLeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheVal 319
Db 812 GATGTTATCAACATGAGCCTTGGCGACCTACTGTTCTACAGCGCTGAAACAGTCGTT 871
QY 320 AspLysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAsp 339
Db 872 GACAAAGCC-----GTTTCCAGCGGTATGCTGCTGCTGCCGCGGAAACGAA 922
QY 340 GlyPro-----LeuTyrGlyThrLeuAsnProAlaAspGlnMetAspValIleGly 357
Db 923 GGTTCATCCGGAAGCACAGACGACGCTACCTGCGCAAAATATCTTCTACTATTGCA 982
QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377

```

1213 983 GTAGTGGTGAACAGCAGCAACCAAGAGCTTCTTCCAGCGAGGTCTT----- 1036
1214 378 TTPGluLeuProGlyGlyTyArgMetLysProAspIleValThrTyArgAlaGly 397
1215 1037 ---GAGCTT-----GATGTGATGGCTCTGCGCGTCC 1066
1216 398 ValArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaLaser 417
1217 1067 ATCCAAAGCAGCTCTCTGGAGGCACTTACGGCGCTTATACGGAACGCTCCATGGCGACT 1126
1218 418 ProValAlaGlyAlaValThrLeuValSerThrValGlnLysArgGluLeuVal 437
1219 1127 CTTCTACTATCGGAGCAGCAGCGTTAATCTTCT-----AAGCAGCGACTTGG 1177
1220 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn 457
1221 1178 ACAAGCGCGCAGCTGATCGTTAGAACGACGTCACATATCTTGA-----AAC 1231
1222 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuArgAla-TyrGlnIleLeuAsnSe 477
1223 1232 TCTTCTACTATGGAAGAGGTTAATCAAGCTACAGCAGCTGCACAA-----AATAG 1285
1224 477 rTyLysProGlnAlaSerLeuSerProSerTyIleAspLeuThrGluCysProTyIle 497
1225 1286 TAAAGAGCAGCGTT-----CCTCATACCTGCTCTTTTATTTGTCAGCAT-- 1334
1226 497 tTrpProTyCysSerGlnProIleTyArgGlyMetProThrValValAsnValTh 517
1227 1335 -----CCTGATGTTCCGCGCGCATTC 1354
1228 517 rIleLeuAsnGlyMet-----GlyValThrGlyArgIle 528
1229 1355 TCTTCTTCTCCGATGTTGAATCGTTCCATGATGACGAGTGGCTGCTCTGAAATC 1414
1230 528 eValAspLysProAspTrpGlnProTyLeuProGlnAsnGly 542
1231 1415 TTCACAGCAGCGAGATCAACCTGCTCAGCCCGCTCAGCGC 1457

```

RESULT 4

```

US-08-322-677A-6
; Sequence 6 Application US/08322677A
; Patent No. 567272
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Chanchal K.
; APPLICANT: Burns, Michael E.
; APPLICANT: DiGiulio, David N.
; APPLICANT: Getty, Edward E.
; APPLICANT: Hartshorn, Richard T.
; APPLICANT: Willey, Alan D.
; APPLICANT: Brode, Philip F., III
; APPLICANT: Barnett, Bobby L.
; APPLICANT: Rubingh, Donn N.
; TITLE OF INVENTION: Bleaching Compositions Comprising Protease Enzymes
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East River Road
; CITY: Cincinnati
; STATE: OH
; COUNTRY: USA
; ZIP: 45253-8707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322.677A
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Zerby, Kim William

```

```

; REGISTRATION NUMBER: 32,323
; REFERENCE/DOCKET NUMBER: 5041R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-2885
; TELEFAX: (513) 627-0318
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1497 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-322-677A-6

```

```

Alignment Scores:
Pred. No.: 8,39e-23 Length: 1497
Score: 331.00 Matches: 124
Percent Similarity: 40.40% Conservative: 59
Best Local Similarity: 27.37% Mismatches: 146
Query Match: 5.89% Indels: 124
DB: 1 Gaps: 19

```

```

US-09-830-837-6 (1-1052) x US-08-322-677A-6 (1-1497)
QY 8 LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGlyLys 27
DB 45 ATAATCTGCTATTGTTGTTTCTGCAAAATGAAAAAGAGAGAGGATAAAGAGTGAGAGGC 104
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
DB 105 AAAAAGATGATGATGATGTTGCTGTTCTGCTTTAGCG-----TTA 143
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
DB 144 ATCTTACGATGCGTTCGCGACACATCTCTCCCGAGCGAGGAAATCAACAGCG 203
QY 53 GluTyGluTyIleValAlaPheAsnGlyTyPheThrAlaLysAlaArgAsnSerPhe 72
DB 204 GAAAGAAATATATGTGCGGTT-----AACACACATGAGCAGC 245
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsn 92
DB 246 ATGAGCGCGCTGAAGAAAGATGTCATT----- 275
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnLysGluLysGlnLysAla 112
DB 276 -----TCTGAAAAAGGC 287
QY 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArgLys 131
DB 288 GGGAAAGTGCAAAAGCAATTCAAATATGTAGACGACGCTTCAGTCACATTTAAACGAAAA 347
QY 132 ValPheArgSerLeuLysTyArgAlaGluSerAspProThrValProCysAsnGluThrArg 151
DB 348 GCTGTAAAGAAATTTGAAA-----AAGACCGAGCGCTCGCT----- 383
QY 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySer 171
DB 383 ----- 383
QY 172 GlyPheThrPheIleAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArg 191
DB 384 -----TACGTTGAAGAGATCAGTACGATCGGTCGCGAGTCCGTCCTTAC 434
QY 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyThrGlyAlaAsn 211
DB 435 GCGGTATACAAA---ATTAAAGCCCTGCTCTGCAAGCTCTCAAGGCTACACATTTAAAGCAAT 491
QY 212 ValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
DB 492 GTTAAAGTAGCGTTATCGACAGCGGTATGATTTCTCTCATCCCTGATTTAAAGGTAGCA 551
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245

```


Qy	192	GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrThrGlyAlaAsn	211
Db	435	GGGTATCAACA---ATTAAGAGCCCTGCTGCACTCTCAAGCTACACTGGATCAAT	491
Qy	212	ValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys	229
Db	492	GTTAAGTAGCGGTATTCGACAGCGGTATCGATTCTCTCATCTGATTTAAAGGTAGCA	551
Qy	230	-----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp	245
Db	552	AGCGGAGCCAGCATGTTCTTCTGAAACAAAT-----CCTTTCCAGAC	596
Qy	246	GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu	262
Db	597	AACAACCTCTCAACGAACCTACGTTGCGGCACAGTTGCGGCTTAATACTCAATCGGT	656
Qy	263	CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln	282
Db	657	GTATTAGGCGTTGCGCAACCGCATCACTTTACGTCGTAAAGTCTTCGGTGCAGCGT	716
Qy	283	ValSerTyThrSerTrpPheLeuAspAlaPheAsnTyTrAlaIleLeuLysIleAsp	302
Db	717	TCCGGCCAAATACAGCTGGATCATTAACGGAATCGAGTGGCGATCGCAACAATATGGAC	776
Qy	303	ValLeuAsnLeuSerIleGlyGlyProAspPheMetasp-----HisProPheValasp	320
Db	777	GTATTAAACATGAGCCTCGCGGACCTTCTGTGTTCTGCTGTTTAAAGGGCAGTTGAT	836
Qy	321	LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly	340
Db	837	AAAGCC-----GTTGCATCGCGCGTCGTAAGTCGTGCGGACCGCGTAAAGAGGC	887
Qy	341	Pro-----LeuTyGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal	358
Db	888	ACTTCGGCAGCTCAAGCACAGTGGCTACCTCGGTGGTAATACCTCTCTCATTTGAGTA	947
Qy	359	GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrp	378
Db	948	GGCGCTGTTGACAGCAGCACCAAGAGCATCTTCTCAAGCGTAGGACCT-----	998
Qy	379	GluLeuProGlyGlyTyArgGlyArgMetLysProAspIleValThrThrGlyAlaGlyVal	398
Db	999	GAGCTT-----GATGTCATGGCACCTGCGGTATCTATC	1031
Qy	399	ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla	416
Db	1032	CRAAGCAGCGTCTCTGGAAACAAATACGGG-----GCGTACAACCGGTACGTCAATGGCA	1085
Qy	417	SerProValValAlaGlyAlaValThrLeuLeuValSer	429
Db	1086	TTCTCCAGCTTCCGCGAGCGGCTGCTGTGATCTTTCT	1124

RESULT 7

RESOLI /
US-08-848-793-6
; Sequence 6, Application US/08848793
; Patent No. 6066611
; GENERAL INVENTION:
; APPLICANT: Ghosh, Chanchal K.
; APPLICANT: Burns, Michael E.
; APPLICANT: DiGiulio, David N.
; APPLICANT: Getty, Edward E.
; APPLICANT: Hartshorn, Richard T.
; APPLICANT: Willey, Alan D.
; APPLICANT: Brode, Philip F., III
; APPLICANT: Barnett, Bobby L.
; APPLICANT: Rubingh, Donn N.
; TITLE OF INVENTION: Bleaching Compositions
; TITLE OF INVENTION: Protease Enzymes
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East River Road
; CITY: Cincinnati

383 ----- 383
 172 GlyPheTrpHisAlaThrGlyArgHisSerArgArgLeuLeuArgAlaIleProArg 191
 384 -----TACGTTGAAGAGATACGATGACATGCGTACGCGGAGTCCGTCGCTTAC 434
 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrGlyAlaAsn 211
 435 GCGGTATCACAA---ATTAAAGCCCGCTCTGCTGCTCTCAAGGCTACATGATCAAT 491
 212 ValArgValAlaValPheAspThrGlyLeuSerGlyHisProHisPheLys----- 229
 492 GTTAAAGTAGCGGTATCGACAGCGGTATGCTCTCTCATCTGATTTAAAGGTAGCA 551
 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
 552 AGCGGAGCGATGCTCTCTTGAACAAAT-----CCTTCCCAAGAC 596
 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
 597 ACAACTCTCAGGAACCTCACGTTGCCGACAGTTCGGCTCTTAATACTCAATCGGT 656
 263 CysGlnGlyPheAlaProAspAlaLeuHisIlePheArgValPheThrAsnGln 282
 657 GTATTAGCGTTCGCCAGCCATCATCTTACGCTGTAAGTCTCGTGTGACGGT 716
 283 ValSerThrTrpPheLeuAspAlaPheAsnTrpAlaIleLeuLysLysIleAsp 302
 717 TCGGCCAATACAGCTGGATCATTAACGGAATCGAGTGGCGATCGCAACAATATGGAC 776
 303 ValLeuAsnSerIleGlyProAspPheMetAsp-----HisProPheValAsp 320
 777 GTATTAACTAGAGCTTCGGCGGACCTTCGTGCTGCTGCTTTAAAGCGCGATGAT 836
 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
 837 AAAGCC-----GTTGCATCGGGGCTGTAGTGTGGCGGAGCGGATACGAAGC 887
 341 Pro-----LeuTyGlyThrLeuAsnProAlaAspGlnMetAspValIleGlyVal 358
 888 ACTTCGGCAGCTCAAGCACAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 947
 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrTrp 378
 948 GCGCTGTTGACAGCAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT----- 998
 379 GluLeuProGlyGlyTyGlyArgMetLysProAspIleValThrTyGlyAlaGlyVal 398
 999 GAGCTT-----GATGTCATGGCAGCTGCGCTATCTATC 1031
 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
 1032 CAAAGCAGCTTCTCGAACAACATACGG-----GCGTACAGCGTACGTCATGCA 1085
 417 SerProValValAlaGlyAlaValThrLeuValSer 429
 1086 TCTCGCAGCTTGGCGGCGCTGCTTGTATCTTCT 1124

RESULT 8
 US-09-445-270-1
 ; Sequence 1, Application US/09445270
 ; Patent No. 6369011
 ; GENERAL INFORMATION:
 ; APPLICANT: Rai, Saroj
 ; APPLICANT: Correa, Paul
 ; APPLICANT: Zhu, Yong
 ; APPLICANT: Graycar, Thomas
 ; APPLICANT: Bott, Richard
 ; TITLE OF INVENTION: Protease Enzymes for Tough Cleaning and
 ; COMPOSITIONS INCORPORATING SAME
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: The Procter & Gamble Company
 STREET: 11810 East Miami River Road
 CITY: Ross
 STATE: OH
 COUNTRY: USA
 ZIP: 45061
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/445,270
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Bolam, Brian M.
 REGISTRATION NUMBER: 37,513
 REFERENCE/DOCKET NUMBER: 1234
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 513/627-2457
 TELEFAX: 513/627-0318
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1497 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-09-445-270-1

Alignment Scores:
 Pred. No.: 8,39e-23 Length: 1497
 Score: 331.00 Matches: 124
 Percent Similarity: 40.40% Conservative: 59
 Best Local Similarity: 27.37% Mismatches: 146
 Query Match: 5.89% Indels: 124
 DB: 4 Gaps: 19

US-09-830-837-6 (1-1052) x US-09-445-270-1 (1-1497)

QY 8 LeuLeuLeuValValLeuLeuLeuCysGlyLysHisLeuGlyAspArgLeuLys 27
 DB 45 ATAATCTCTCTATTGTTATCTGCAAAATGAAAAAGAGAGGATAAAGTGAGAGC 104
 QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
 DB 105 AAAAAAGTATGATCAGTTGCTGTTGCTTTAGCG-----TTA 143
 QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
 DB 144 ATCTTTACGATGCGGTTTCGGCAGCATCTCTGCCAGCGGAGGAAATCAACGGG 203
 QY 53 GluTyGluTyIleValAlaPheAsnGlyTyPheThrAlaLysAlaArgAsnSerPhe 72
 DB 204 GAAAGAATATATTGCGGGTTT-----AACACAGACATGAGCAGC 245
 QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsn 92
 DB 246 ATGACGCGCGCTAAGAGAAGATGTCATT----- 275
 QY 93 ProSerSerAspTyProSerAspPheGluValIleGlnIleLysGlnLysAla 112
 DB 276 -----TCTGAAAAAGGC 287
 QY 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArgLys 131
 DB 288 GGAAGAGTCAAGAGCAATTCAAATATGTAGACGAGCTTCATCATTAACCAANA 347
 QY 132 ValPheArgSerLeuLysTyIleGluSerAspProThrValProCysAsnGluThrArg 151
 DB 348 GCTGTAAAGAAATTGAAA-----AAAGACCCGAGCGTCGCT----- 383

QY 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySer 171
Db 383 -----
QY 172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArg 191
Db 384 -----TACGTTGAAGAAGATCAGTAGACATCGCGPACCGGAGTCCGCGCTTAC 434
QY 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsn 211
Db 435 GCGGTATACAA--ATTAAAGCCCTGCTCTGCACCTCTCAAGCTACACTGATCAAT 491
QY 212 ValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 492 GTTAAAGTAGCGTTATCGACAGCGTATCGATCTCTCATCTGATTTAAAGGTAGCA 551
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
Db 552 ACGGGAGCCAGCATGGTCTCTGAAACAAAT-----CCTTTCCAAGAC 596
QY 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 597 AACAACTCTCAGGAACTCAGCTGCGGACAGTTCGGCTCTTAATACTCAATCGGT 656
QY 263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
Db 657 GTATTAGGGCTTCGCGCAAGCGCATCTTACGCTGTAAAGTCTCTCGGTGCTGCGGT 716
QY 283 ValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAsp 302
Db 717 TCGGGCAATACAGCTGGATCATTAACGGAATCAGTGGCGGATCGCAACAATATGGAC 776
QY 303 ValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAsp 320
Db 777 GTATTAAATAGAGCTTCGCGGACCTCTGGTCTGCTGCTTTAAAGCGGCGAGTGTAT 836
QY 321 LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly 340
Db 837 AAAGCC-----GTTCATCCGGCTGCTAGTCTGTTGCGGACGCGGTAACGAAGGC 887
QY 341 Pro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
Db 888 ACTTCGGCAGCTCAAGACAGTGGCTACCCCTGGTAAATACCCCTCTGTCATTCGAGTA 947
QY 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrp 378
Db 948 GGGCTGTGACAGCAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT----- 998
QY 379 GluLeuProGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyVal 398
Db 999 GAGCTT-----GATGTCATGGCACCTGGCGTATCTATC 1031
QY 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1032 CAAGCAGCGTCTCTGGAACAAATACGG-----CGCTACACAGGTAGTCAATGGCA 1085
QY 417 SerProValValAlaGlyAlaValThrLeuLeuValSer 429
Db 1086 TCTCCGACGCTTCGCGAGCGGCTGCTTTGATCTTCT 1124

RESULT 9

US-09-178-155-1
; Sequence 1, Application US/09178155
; Patent No. 6312936
; GENERAL INFORMATION:
; APPLICANT: Poulos, Ayrookaran J.
; APPLICANT: Schellenberger, Volker
; APPLICANT: Kellis, Jr., James T.
; APPLICANT: Paech, Christian
; APPLICANT: Nachtern, Joanne
; APPLICANT: Naki, Donald P.
; TITLE OF INVENTION: Multiply-Substituted Protease Variants
; FILE REFERENCE: GC502-2

; CURRENT APPLICATION NUMBER: US/09/178,155
; CURRENT FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 08/956,323
; EARLIER FILING DATE: 1997-10-23
; EARLIER APPLICATION NUMBER: 08/956,564
; EARLIER FILING DATE: 1997-10-23
; EARLIER APPLICATION NUMBER: 08/956,324
; EARLIER FILING DATE: 1997-10-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: B. amyloliquefaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (96)...(1245)
US-09-178-155-1

Alignment Scores:

Pred. No.: 8,54e-23 Length: 1513
Score: 331.00 Matches: 124
Percent Similarity: 40.40% Conservative: 59
Best Local Similarity: 27.37% Mismatches: 146
Query Match: 5.89% Indels: 124
DB: 4 Gaps: 19

US-09-830-837-6 (1-1052) x US-09-178-155-1 (1-1513)

QY 8 LeuLeuLeuValValLeuLeuCysGlyLysHisLeuGlyAspArgLeuGluLys 27
Db 45 ATAATCTGTCTATTGGTTATCTGCAATGAAAAAGAGAGGATAAAGAGTGAGAGGC 104
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 105 AAAAAAGTATGATCATGTTGCTGTTGCTTTAGCG-----TTA 143
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 144 ATCTTTACGATGGCTTCGCGACGACATCTCTGCCAGCGGCGGAGGAAATCAACGGG 203
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 204 GAAAGAAATATATATGTCGGGTT-----AAACAGACAATGAGCAGC 245
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIlelleProArgAsn 92
Db 246 ATGAGCGCGCTAAGAAAGAAAGATGTCATT----- 275
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 276 -----TCTGAAAAGGC 287
QY 113 GlyLeuLeuThrLeuGlu-----AspHisProAsnIleLysArgValThrProGlnArgLys 131
Db 288 GCGAAGATGCAAAAGCAATATCAATATATAGAGCGAGCTTCAGCTACATTAACGAAAA 347
QY 132 ValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArg 151
Db 348 GCTGTAAAGAAATTGAAA-----AAAGACCGGAGCGTCT 383
QY 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySer 171
Db 383 ----- 383
QY 172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArg 191
Db 384 -----TACGTTGAAGAAGATCAGTAGACATCGGTACGCGAGTCCGTCCTTAC 434
QY 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsn 211
Db 435 GCGGTATACAA--ATTAAAGCCCTGCTCTGCTGCTCTCAAGGCTACACTGATCAAT 491

QY 212 ValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 492 GTTAAAGTAGGGTTATCGACAGCGGTATGATCTTCATCTCTGATTTAAAGGTAGCA 551
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
Db 552 AGCGGAGCAGCATGGTCTCTTCTCAACAAAT-----CCTTCCCAAGAC 596
QY 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 597 AACAACTCTACGGAACCTACGTTCCGCGCACAGTTGCGGCTCTTAATAACTCAATCGGT 656
QY 263 CysGlnGlyPheAlaProAspAlaGluLeuHisPheArgValPheThrAsnAsnGln 282
Db 657 GTATTAGCGTTCGCCCAAGCGCATCATCTTACGCTGTAAGTTCTCGGTGTGACGCT 716
QY 283 ValSerTrpThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAsp 302
Db 717 TCGGCGCAATACAGCTGGATCATTAACGAATCGAGTGGCGCATCGCAACAATATGGAC 776
QY 303 ValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAsp 320
Db 777 GTATTACATGAGCTTCGCGGACCTTCTGCTTCTGCTTTAAAGCGCGAGTGTAT 836
QY 321 LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly 340
Db 837 AAGGCC-----GTTGATCTCGGCGTGTAGTCTGCGGCGAGCGGTACGAGGC 887
QY 341 Pro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
Db 888 ACTTCGCGAGCTCAAGCACAGTGGCTTACCCTGGTAAATACCTCTGTCATTCGAGTA 947
QY 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrTrp 378
Db 948 GCGCTGTGTACAGCACAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT----- 998
QY 379 GluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyVal 398
Db 999 GAGCTT-----GATGTCATGGCACCTGGCTATCTATC 1031
QY 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1032 CAAAGCAGCGTCTCTGGAACAATAACGCG-----GCGTACAAGGTACGTCAATGGCA 1085
QY 417 SerProValAlaGlyAlaValThrLeuValSer 429
Db 1086 TCTCGGACGTTGCGGAGCGGCTGCTTGATCTTCT 1124

RESULT 10
US-09-000-016-1
Sequence 1, Application US/09000016
Patent No. 6143541
GENERAL INFORMATION:
APPLICANT: Akira ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998

CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE: Streptomyces viridosporus
STRAIN: A-914
ORGANISM: Streptomyces antibioticus
ORIGINAL SOURCE:
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores:
Pred. No.: 3,14e-22 Length: 2809
Score: 330.00 Matches: 169
Percent Similarity: 36.98% Conservative: 81
Best Local Similarity: 25.00% Mismatches: 258
Query Match: 5.88% Indels: 169
DB: 3 Gaps: 28
US-09-830-837-6 (1-1052) x US-09-000-016-1 (1-2809)
QY 125 ArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSer----- 141
Db 745 CGTCGGCTACCGGGCGCCGCCAGCGGCCGCCAGCGCGCGGAGGACTCCGCGAGCGGGGAACT 804
QY 142 -----AspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSer 159
Db 805 CCGCGGAGCCCTGACGTCCCT-----GAACCGGAGCGCGGTCCG 843
QY 160 ArgProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheThrHisAlaThrGlyArg 179
Db 844 GACCCCGCACGA-----GGACGCGTCCGAGCTGTGGGACGC 879
QY 180 HisSerSerArgArgLeuLeuArg----- 187
Db 880 CGTCACCAACCGGACCGGACCGGCGGATCGCGCATCGCCAGCTGTGGAGGGGTCCG 939
QY 188 ---AlaIleProArgGlnValAlaGlnThr-LeuGlnAlaAspValLeuTrpGlnMetG1 206
Db 940 CAGGCGCCCTCGACACGTCCTCGGCGAGATCGGCGCCGCCAACGCGGTGTCCGCGCG 999
QY 206 YTrpThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPr 226
Db 1000 CTACGACGCGCAAGGGGTGAAGATCGCGTCTCTGACACCGGTGTGACACGAGCATCC 1059
QY 226 ohisPhelys---AsaValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAs 245
Db 1060 GACCTGTGAGGCGGCGGTGACCGCGTCCCAAGACTTCCCGCGCGCGCGCGCGCGA 1119

Db 1194 TCCTTCTACTATGGAAGGCGTGATCAACGTACAGCGCGCAGCTCAG 1241

RESULT 14

5472855-1

Patent No. 5472855

APPLICANT: CARTER, PAUL J.; WELLS, JAMES A.

TITLE OF INVENTION: SUBSTRATE ASSISTED CATALYSIS

NUMBER OF SEQUENCES: 31

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/287,964

FILING DATE: 22-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 90,902

FILING DATE: 12-JUL-1993

APPLICATION NUMBER: 823,039

FILING DATE: 14-JAN-1992

APPLICATION NUMBER: 35,652

FILING DATE: 06-APR-1987

APPLICATION NUMBER: 334,081

FILING DATE: 04-APR-1989

APPLICATION NUMBER: 127,134

FILING DATE: 01-DEC-1987

APPLICATION NUMBER: 846,627

FILING DATE: 01-APR-1986

APPLICATION NUMBER: 858,594

FILING DATE: 30-APR-1986

APPLICATION NUMBER: 614,612

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,615

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,617

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,491

FILING DATE: 29-MAY-1984

SEQ ID NO:1:

LENGTH: 1497

5472855-1

Alignment Scores:

Pred. No.:	1,48e-22	Length:	1497
Score:	328.50	Matches:	130
Percent Similarity:	40.32%	Conservative:	70
Best Local Similarity:	26.21%	Mismatches:	169
Query Match:	5.85%	Indels:	127
DB:	6	Gaps:	21

US-09-830-837-6 (1-1052) x 5472855-1 (1-1497)

QY	8	LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys	27
Db	45	ATAATCTGTCTATTGTTTCTGCAATGAAAAAGGAGAGGATAAAGAGTGAGAGGC	104
QY	28	LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu	41
Db	105	AAAAAGTATGATCATGTTGCTGTTGCTTTAGCG-----TTA	143
QY	42	ThrLeuLysValGluPheSerSerThrValVal-----	52
Db	144	ATCTTTACGATGGCGTTGCGGCACACATCTCTGCCCGCGCGCAATCAACGGG	203
QY	53	GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe	72
Db	204	GAAGAAGATATATTCGCGGTTT-----AAACAGACAATGAGCAGC	245
QY	73	IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn	92
Db	246	ATGAGCGCGCTAAGAAGAAAGATGTCATT-----TCTGAA	281
QY	93	ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla	112
Db	282	AAAGCGGGAAGTGAAGCAATTCATATGATAGC-----GCA	323
QY	113	GlyLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal	132

Db	324	GCTTCAGCTACATTAAACGAA-----	344
QY	133	PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTrp	152
Db	345	-----AAAGCTGTAAGAAATTTGAAAAAGACCCGAGCGTCGCT-----	383
QY	153	SerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySerGly	172
Db	383	-----	383
QY	173	PheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArgGln	192
Db	384	-----TACGTTGAAGAAGATCAGTAGCACATCGGTACGCGAGTCGCTTACGGC	437
QY	193	ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnVal	212
Db	438	GTATCAAA---ATTAAAGCCCTGCTCTCATCTCTCAAGCTACATCGATCAATGTT	494
QY	213	ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys-----	229
Db	495	AAAGTAGCGGTTATCGACACGCGTATCGATTCTTCATCTGATTTAAAGGTAGCAGC	554
QY	230	-----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGly	246
Db	555	GGAGCCAGCATGTTCTCTGAACAAAT-----CCTTTCCAAGACAAC	599
QY	247	LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys	263
Db	600	AACTCTCAGGAACCTCAGTTCGCGCACGTCTTCTTAATACTCAATCGGCGTA	659
QY	264	GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal	283
Db	660	TTAGCGGTTGCGCAAGCGCATCATCTTACGCTGTAAGAGTTCTCGGTGCTGAGCGTCC	719
QY	284	SerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal	303
Db	720	GGCCAATACAGCTGGATCATTAACGGAATCAGTGGCGCATCGCAACAATATGAGCGTT	779
QY	304	LeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheValAspLys	321
Db	780	ATTAACATGAGCTCGCGGAGCCTCTGCTGCTGCTTAAAGCGCGCATGTATGATAA	839
QY	322	ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro	341
Db	840	GCC-----GTGATCCGCGCGTGTAGTGTTCGCGCAGCGCGTAAAGAGGCACT	890
QY	342	-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly	359
Db	891	TCCGGCAGCTCAAGCACAGTGGGTACCTTGTGTAATAACCTTCTGTCATTGCAAGT	950
QY	360	GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu	379
Db	951	GCTGTTGACAGCAGCAACAAAGAGCATCTTCTCAAGCTAGGACCT-----GAG	1001
QY	380	LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArg	399
Db	1002	CTT-----GATGTCATGCGCACCTGCGGTATCTATCTCA	1034
QY	400	GlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer	417
Db	1035	AGCACGCTCTCTGGAACAAATACGGG-----CGGTACACACGCGTACGTAATGGCATCT	1088
QY	418	ProValValAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal	437
Db	1089	CCGCGCTGCGCGGCGGCTGCTTGTATCTTTCT-----AAGCACCCGCACTGG	1139
QY	438	AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn	457
Db	1140	ACAAACATCAAGTCCGCGCAGCTTTAGAAAACACCACCTACAAATCTGGT-----GAT	1193
QY	458	MetPheGlnGlnHisGlyLysLeuAspLeuLeuArgAlaTyrGln	473

Db 1194 TCTTTACTATGAAAAGGGGTGATCAACGTACAGCGGCGAGCTCAG 1241

RESULT 15

US-08-069-863-1
; Sequence 1, Application US/08069863
; Patent No. 5470733
; GENERAL INFORMATION:
; APPLICANT: BRYAN, Philip N
; APPLICANT: ALEXANDER, Patrick
; APPLICANT: STRAUSSBERG, Susan L
; TITLE OF INVENTION: CALCIUM FREE SUBTILISIN MUTANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/069.863
; FILING DATE: 01-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 028755-016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 450..1599
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 772..1599
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 450
; OTHER INFORMATION: /note= "Amino Acid Val at position
; OTHER INFORMATION: 450 is fMet."
US-08-069-863-1

Alignment Scores:
Pred. No.: 2,18e-22 Length: 1868
Score: 328.50 Matches: 130
Percent Similarity: 40.32% Conservative: 70
Best Local Similarity: 26.21% Mismatches: 169
Query Match: 5.85% Indels: 127
Gaps: 21

US-09-830-837-6 (1-1052) x US-08-069-863-1 (1-1868)

Qy 8 LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGlyLys 27
Db 399 ATAAATGCTGCTATTGTTATTCGTAATGAAAAGGAGGAGGATGAAGAGTGAAGGC 458
Qy 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 459 AAAAAGTATGGATGATGCTGCTTTGCTTAGCG-----TTA 497

Qy 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 498 ATCTTTACGATGGCTTCCGACGACACCTCTCCAGCGCGGAGGAAATCAACGGG 557
Qy 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaAsnSerPhe 72
Db 558 GAAAGAAATATATTCGCGTT-----AAACAGACAATGAGCAGC 599
Qy 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsn 92
Db 600 ATGACGCCCGCTAAGAGAAAGATGTCATT-----TCGAA 635
Qy 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 636 AAGCGGGGAAAGTCAAAAGCAATTAATATGTAGC-----GCA 677
Qy 113 GlyLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
Db 678 GCTTCAGCTACATTAACGAA----- 698
Qy 133 PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
Db 699 ---AAAGCTGTAAAGAATTTGAAAAGACCCGAGCTCGCT----- 737
Qy 153 SerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySerGly 172
Db 737 ----- 737
Qy 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGln 192
Db 738 -----TACGTTGAAGAAGATCACGTAGCACATCGTACGCGAGTCCGCTGCTTACGGC 791
Qy 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnVal 212
Db 792 GTATCACA---ATTAAAGCCCTCTCTGCACCTCTCAAGGCTACATGGATCAATGTT 848
Qy 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 849 AAGTAGCGGTATTCAGACGCGTATCCATTCCTCCTCACTTCATTAAGGTAGCAGGC 908
Qy 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGly 246
Db 909 GGAGCCAGCATGGTCTCTCTGAAACAAAT-----CCTTCCAGACACAC 953
Qy 247 LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263
Db 954 AACTCTCAGGAACTCACGTCGCGCACAGTTCGCGCTCTTAATAACTCAATCGGTGTA 1013
Qy 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal 283
Db 1014 TTAGCGGTTGCGCAAGCGCATCACTTTACGCTGTTAAAGTTCTCGGTGCTGAGCGTCC 1073
Qy 284 SerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal 303
Db 1074 GGCCAATACAGCTGGATCATTAACGGAATCGAGTGGCGATCGCAACAATATGACGTT 1133
Qy 304 LeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAspLys 321
Db 1134 ATTAACATGAGCTCGCGGACCTTCTGTTCTGCTTAAAGCGCGCAGTTGATATA 1193
Qy 322 ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341
Db 1194 GCC-----GTTGCATCCGCGCTGCTGCGCGCGCGGTAACGAGGACACT 1244
Qy 342 -----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
Db 1245 TCCGCGACGCTCAAGCACACATGGCTACCTGGTAAATPACCTTCTGTCTATTCAGTAGGC 1304
Qy 360 GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu 379
Db 1305 GCTGTTGACAGCAGCAACCAAGAGCATCTTTCTCAAGCTAGGACCT-----GAG 1355
Qy 380 LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrThrGlyAlaGlyValArg 399

Db 1356 CIT
QY 400 GlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
Db 1389 AGCAGCTTCCTGGAAACAATACGG-----GCGTACACGGTACGTCATGGCATCT 1442
QY 418 ProValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
Db 1443 CGCAGCTTCCGGAGCGCTGTTGATCTCTTCT-----AAGCACCCGAACTGG 1493
QY 438 AsnProAlaSerMetLysGlnAlaLeuLeuAlaSerAlaArgLeuProGlyValAsn 457
Db 1494 ACAACACACTCAAGTCGCGAGCGCTGTTAGAAAAACACCACTACAAAACCTGGT-----GAT 1547
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGln 473
Db 1548 TCTTCTACTATGGAAAGGGCTGATCAACGTACAGCGCGACGCTCAG 1595

RESULT 16

US-08-309-069-1
; Sequence 1, Application US/08309069
; Patent No. 5567601
; GENERAL INFORMATION:
; APPLICANT: BRYAN, Philip N
; APPLICANT: ALEXANDER, Patrick
; APPLICANT: STRAUSBERG, Susan L
; TITLE OF INVENTION: CALCIUM FREE SUBTILISIN MUTANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,069
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 028755-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 450..1599
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 772..1599
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 450
; OTHER INFORMATION: /note= "Amino Acid Val at position 450 is fMet."
US-08-309-069-1

Alignment Scores:

Pred. No.: 2,18e-22 Length: 1868
Score: 328.50 Matches: 130
Percent Similarity: 40.32% Conservative: 70
Best Local Similarity: 26.21% Mismatches: 169
Query Match: 5.85% Indels: 127
DB: 1 Gaps: 21
US-09-830-837-6 (1-1052) x US-08-309-069-1 (1-1868)
QY 8 LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys 27
Db 399 ATAATCTGCTATGTGTTATCTGCAAAATCAAAAGAGAGAGATAAGAGTGAGAGGC 458
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 459 AAAAAAGTATGGATCATGTTGCTGTTTGTCTTTAGCG-----TTA 497
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 498 ATCTTTACGATGGCGTTTCGGCAGCATCTCTGCCAGCGCGAGGAAATCAACAGGG 557
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 558 GAAAGAGAAATATATTCTCGGTTT-----AAACAGACAATGAGCAGC 599
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTyrArgIleIleProArgAsnAsn 92
Db 600 ATGAGCGCGCTAAGAGAAAGATGTCATT-----TCGAA 635
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 636 AAAGCGGGAAGTGAAGAAAGCAATCAAAATATGTAGAC-----GCA 677
QY 113 GlyLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
Db 678 GCTTCAGTACATTAAACGAA----- 698
QY 133 PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
Db 699 ---AAAGCTGTAAACAATGAAAAAGACCCGAGGCTCGCT----- 737
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySerGly 172
Db 737 ----- 737
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGln 192
Db 738 -----TACGTTCAAGAAGATCAGTAGCACATGCGTACGCGCAGTCCGCTTACGGC 791
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnVal 212
Db 792 GTATCACA---ATTAAAGCCCTGCTCTGCACTCTCAAGGCTACACTGGATCAATGTT 848
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 849 AAAGTAGCGGTTATCGACAGCGGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 908
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGly 246
Db 909 GGAGCCAGCATGGTCT 953
QY 247 LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263
Db 954 AACTCTCAGGAACTCAGCTTCCGCGCAGAGTTCGGCTCTTAATCACTCAATCGGTGTA 1013
QY 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal 283
Db 1014 TTAGCGGTTGCGCAAGCGCATCATTACGCTGTAAAGATCTCTCGTGTCTCGGCTTCC 1073
QY 284 SerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal 303
Db 1074 GGCCAATACAGCTGGATCATTAACGGAATCGAGTGGCGCATCGCAAAACAATATGGACGT 1133


```
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
849 AAAGTAGCGTATCCACACGGTATCGATTCTCTCATCCTGATTAAGGTAGCAGC 908
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspGly 246
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
909 GGAGCAGCAGTGTCTCTCAACAAT-----CCTTCCAAAGACAAC 953
QY 247 LeuGlyHisGlyThrPheValAlaGlyValLleAlaSerMetArgGlu-----Cys 263
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
954 AACTCTCAGGAAGTACGTCGCGCACAGTTCGCGCTCTTAATCAATCAATCGGTGA 1013
QY 264 GlnGlyPheAlaProAspAlaGluLeuHisLlePheArgValPheThrAsnAsnGlnVal 283
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1014 TTAGCGCTGCGCAACGACATCTTACCGTGTAAAGTTCTCGGTGCTGCGGTTC 1073
QY 284 SerThrThrSerTrpPheLeuAspAlaPheAsnTrpAlaLleLeuLysLysLysVal 303
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1074 GGCAATACAGCTGGATCATTAACGAATCGAGTGGCGATCCAAACAATATGACGT 1133
QY 304 LeuAsnLeuSerLleGlyGlyProAspPheMetAsp-----HisProPheValAspLys 321
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1134 ATTAACATGAGCTCGCGGACCTCTGCTGCTGCTTAAACCGCGCAGTTGATAA 1193
QY 322 ValTrpGluLeuThrAlaAsnAsnValLleMetValSerAlaLleGlyAsnAspGlyPro 341
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1194 GCC-----GTTGCATCGCGCGTGTAGTGTGCGGCAGCGGTAAAGGCACT 1244
QY 342 -----LeuTrpGlyThrLeuAsnAsnProAlaAspGlnMetAspValLleGlyValGly 359
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1245 TCGCGCAGCTCAAGCAGACATGGCTACCTGCTGTAATACCTCTCTCATTCAGTAGGC 1304
QY 360 GlyIleAspPheGluAsnLleAlaArgPheSerSerArgGlyMetThrThrTrpGlu 379
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1305 GCTGTTGACAGCAACCAAGAGCATCTTCTCAAGCTAGGACCT-----GAG 1355
QY 380 LeuProGlyGlyTrpGlyArgMetLysProAspLleValThrTyrglyAlaGlyValArg 399
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1356 CTT-----GATGTCATGCACCTGGCGTATCTATCCAA 1388
QY 400 GlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1389 AGCAGCTCTCTGGAACAAATACGG-----GCGTACACGGTACGTCAATGGCATCT 1442
QY 418 ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1443 CCGCAGTTCGCGGCGGCTGCTTGTATCTTCT-----AGCACCCGACTGG 1493
QY 438 AsnProAlaSerMetLysGlnAlaLeuLleAlaSerAlaArgArgLeuProGlyValAsn 457
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1494 ACAACACATCAAGTCGCGCAGCTTTAGAAAACACCACCTACAAAACCTTGGT-----GAT 1547
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrgln 473
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1548 TCTTCTACTATGAAAGGCTGATCAACGTACAGCGGCGCAGCTCAG 1595
```

RESULT 18

PCT-US95-05520-1

; Sequence 1. Application PC/TUS9505520

; GENERAL INFORMATION:

; APPLICANT: BRYAN, Phillip N

; APPLICANT: ALEXANDER, Patrick

; APPLICANT: STRAUSBERG, Susan L

; TITLE OF INVENTION: CALCIUM FREE SUBSTITILIN MUTANTS

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05520
; FILING DATE: 28-APR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/309,069
; APPLICATION NUMBER:
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 028758-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 450..1599
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 772..1599
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 450
; OTHER INFORMATION: /note= "Amino Acid Val at position
; OTHER INFORMATION: 450 is fMet."
; PCT-US95-05520-1
```

Alignment Scores:

Pred. No.:	2,18e-22	Length:	1868
Score:	328.50	Matches:	130
Percent Similarity:	40.32%	Conservative:	70
Best Local Similarity:	26.21%	Mismatches:	169
Query Match:	5.85%	Indels:	127
DB:	5	Gaps:	21

US-09-830-837-6 (1-1052) x PCT-US95-05520-1 (1-1868)

```
QY 8 LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys 27
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
399 ATAATCTGCTATTGTTGTTATCTGCAAAATGAAAAAGGAGAGGATAAAGAGTCAGAGGC 458
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
459 AAAAAAGTATGGATCATGTTGCTGTTTGTCTTAGCG-----TTA 497
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
498 ATCTTTACGATGCGGTTCGCACACACATCTCTGCCCGCGCGGAGAAATCAACGGG 557
QY 53 GluTyrgluTyrlleValAlaPheAsnGlyTyrrPheThrAlaLysAlaArgAsnSerPhe 72
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
558 GAAAGAAATATATTGTCGGGTT-----AAACAGACAATGAGCAGC 599
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgillelleProArgAsnAsn 92
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
600 ATGAGCGCGCTAAGAAAGAAAGATGTCATT-----TCTGAA 635
QY 93 ProSerSerAspTyrrProSerAspPheGluValLleGlnLleLysGluLysGlnLysAla 112
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
636 AAAGCGGGAAAGTGCAAAAGCAATTCAAATATGTAGAC-----GCA 677
QY 113 GlyLeuLeuThrLeuGluAspHisProAsnLleLysArgValThrProGlnArgLysVal 132
```


253 ATGAGCGCGCTAAGAAGAAGATGTCATT-----TCTGAA 288
Db
93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
QY ||||| : : :
Db 289 AAGGCGGGAAGTGAAGAAGCAATTCATATGTAGAC-----GCA 330
QY 113 GlyLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
Db ||||| : : :
Db 331 GCTTCAGCTACATTAAACGAA----- 351
QY 133 PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTyr 152
Db : : : : : ||||| : : : : :
Db 352 ---AAGCGTGAAGAAGTGAAGAAGACCGCGCTCGCT----- 390
QY 153 SerGlnLysTyrPheSerArgProLeuArgArgAlaSerLeuSerGlySerGly 172
Db ----- 390
QY 173 PheTyrHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGln 192
Db : : : : : ||||| : : : : :
Db 391 ---TACGTTGAAGAAGATCACGTAGCACATCGGTAGCGCGTCCGTACGGC 444
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTyrGlnMetGlyTyrThrGlyAlaAsnVal 212
Db ||||| : : : : : ||||| : : : : :
Db 445 GTATCAAA---ATTAAAGCCCTGCTCTCAAGGCTACACTGGATCAATGTT 501
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db : : : : : ||||| : : : : :
Db 502 AAGTAGCGGTATTCACACGCGGTATCGATTCTTCATCCTCATTAAGGTAGCAGGC 561
QY 230 -----AsnValLysGluArgThrAsnTyrThrAsnGluArgThrLeuAspAspGly 246
Db ||||| : : : : :
Db 562 GGAGCAGCATGTTCTCTGAACAATA-----CCTTTCCAAGACAAC 606
QY 247 LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263
Db ||||| : : : : : ||||| : : : : :
Db 607 GACTCTCAGGAAGTACAGTTCGCGGACAGTTCGCGCTCTTAATAACCAATCAATCGGTGA 666
QY 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal 283
Db ||||| : : : : : ||||| : : : : :
Db 667 TTAGCGTTCGCGCAAGCGCATCATCTAGCGTGTAAAGTTCTCGTGTGCGGTTC 726
QY 284 SerTyrThrSerTyrPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal 303
Db : : : : : : : : : : : : : : :
Db 727 GGCAATACAGCTGGATCATTAACGAATCGAGTGGCGATCGCAACAATATGCGAGTT 786
QY 304 LeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAspLys 321
Db : : : : : ||||| : : : : :
Db 787 ATTAACATGAGCCTCGCGGACCTCTGTTCTGCTGCTTTAAAGCGCGAGTGTATAA 846
QY 322 ValTyrGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341
Db ||||| : : : : : ||||| : : : : :
Db 847 GCC-----GTTGATCCGCGCTAGTCTGCGGACCGCGTACGAAGGCACT 897
QY 342 -----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
Db ||||| : : : : : ||||| : : : : :
Db 898 TCGGCGAGCTGCTCGAGTGGACTACCTCGCAATACCTCTCTGTCATGTGAGTAGGC 957
QY 360 GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThrTyrGlu 379
Db : : : : : : : : : : : : : : :
Db 958 GCTGTTGACAGCAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT-----GAG 1008
QY 380 LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArg 399
Db ||||| : : : : : ||||| : : : : :
Db 1009 CTT-----GATGTCATGCGACCTGCGCTATCTATCCAA 1041
QY 400 GlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
Db : : : : : ||||| : : : : : ||||| : : : : :
Db 1042 AGCACGCTCCTGGAACAATAACGG-----GCGTACACAGGTACCTCAATGGCATCT 1095
QY 418 ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
Db ||||| : : : : : ||||| : : : : :
Db 1096 CCGCAGCTTGCAGGCGGTGCTTTGATTCTTCT-----AAGCACCGGAACTGG 1146

QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsn 457
Db : : : : : ||||| : : : : :
Db 1147 ACAACACATCAAGTCGCGACAGTTTAGAAAAACACCACTACAAACTTGGT-----GAT 1200
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGln 473
Db ||||| : : : : : ||||| : : : : :
Db 1201 TCTTTCTACTATGGAAGGCTGATCAACGTCAGGCGGCGAGCTCAG 1248

RESULT 20

US-08-398-028B-1
; Sequence 1, Application US/08398028B
; Patent No. 5780285
; GENERAL INFORMATION:
; APPLICANT: Marcus D. Ballinger and James A. Wells
; TITLE OF INVENTION: SUBSTITUTED VARIANTS CAPABLE OF CLEAVING
; TITLE OF INVENTION: SUBSTRATES CONTAINING DIBASIC RESIDUES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/398, 028B
; FILING DATE: 03-Mar-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kubinec, Jeffrey S.
; REGISTRATION NUMBER: 36,575
; REFERENCE/DOCKET NUMBER: P0936
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-8228
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8119 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-398-028B-1

Alignment Scores:

Pred. No.: 2,79e-21 Length: 8119
Score: 328.50 Matches: 130
Percent Similarity: 40.52% Conservative: 71
Best Local Similarity: 26.21% Mismatches: 168
Query Match: 5.85% Indels: 127
DB: 1 Gaps: 21

US-09-830-837-6 (1-1052) x US-08-398-028B-1 (1-8119)

QY 8 LeuLeuLeuValValLeuLeuLysGlyLysLysHisLeuGlyAspArgLeuGluLys 27
Db : : : : : ||||| : : : : :
Db 52 ATAATCTGCTATTGGTTATTCGCAATCAAAAAGAGAGAGATAAAGAGTCAAGGC 111
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db ||||| : : : : : ||||| : : : : :
Db 112 AAAAAGATGATGATCAGTTGCTGTTGCTTAGCG-----TTA 150
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db : : : : : ||||| : : : : :
Db 151 ATCTTTACGATGCGGTTCGCGACGACATCTCTGCCAGCGGCGAGGAATCAACGGG 210
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : |||||

Db 211 GAAAGAAATATATTGTCGGGTTT-----AAACAGACAAATGACGACG 252
 Qy 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn 92
 Db 253 ATGAGCGCGCTAAGAAAGAAAGATGCTATT-----TCGAA 288
 Qy 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
 Db 289 AAGCGCGGAAAGTGCAGAAAGCAATTCMAATATGTAGAC-----GCA 330
 Qy 113 GlyLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
 Db 331 GCTTCAGCTACATTAAACGAA----- 351
 Qy 133 PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
 Db 352 ---AAGCTGTAAAGAAATGAAAGACCGGAGCGTCTGCT----- 390
 Qy 153 SerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySerGly 172
 Db 390 ----- 390
 Qy 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGln 192
 Db 391 -----TACGTTGAAGAAATCAGTACGACATCGCTACGCGAGTCCGCTTACGCG 444
 Qy 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnVal 212
 Db 445 GTATCAAA---ATTAAAGCCCTCTCTGCTGACTCTCAAGGCTACACTGATCAATGTT 501
 Qy 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
 Db 502 AAGTAGCGGTATCGACAGCGGTATCGATTCTCTCATCTGATTAAAGGTAGCAGCG 561
 Qy 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGly 246
 Db 562 GGAGCCAGCATGTTCTCTCTGAAACAAAT-----CCTTCCAGACAAAC 606
 Qy 247 LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263
 Db 607 GACTCTCAGCAACTCAGCTTCGCGCACAGTTCGCGCTCTTAATACTCAATCGGTGTA 666
 Qy 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal 283
 Db 667 TTAGCGGTTCGCCAAGCGCATCACTTACGCTGTAAAGTTCTCGGTGCTGACGTTCC 726
 Qy 284 SerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal 303
 Db 727 GGCAATACAGTGGATCATTAAGGAATCGAGTGGCGATCGCAACAAATATGAGCTT 786
 Qy 304 LeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheValAspLys 321
 Db 787 ATTACATGAGCTCGCGGACCTTCGTCTGCTGCTTTAAAGCGGCGAGTTGATAA 846
 Qy 322 ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341
 Db 847 GCC-----GTTGCATCCGGCTGTAGTCTGTCGGCGAGCGGTACGAAGGACT 897
 Qy 342 -----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
 Db 898 TCGCGACGCTCGCAGACTGAGCTACCCCTGGCAATACCTCTGTCATTGTCAGTAGGC 957
 Qy 360 GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThrTrpGlu 379
 Db 958 GCTGTTGACAGACGACCAACAAAGACGATCTTCTCAAGCGTAGGACCT-----GAG 1008
 Qy 380 LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyValAlaGlyValArg 399
 Db 1009 CTT-----GATGTCATGGCAGCTGCGCTATCTATTCAA 1041
 Qy 400 GlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
 Db 1042 ACCAGCTCTCTGGAACAAATACGCG-----GCTACACGCTGCTCAATGCGATCT 1095

Qy 418 ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
 Db 1096 CCGCAGGTTCGGAGCGGCTGCTTGTATTCTTCT-----AAGCACCAGCTGG 1146
 Qy 438 AsnProAlaSerMetLysGlnAlaLeuLeaSerAlaArgLeuProGlyValAsn 457
 Db 1147 ACAACACACTCAAGTCCGCGACGATTTAGAAAAACACCACTACAAACTTGGT-----GAT 1200
 Qy 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGln 473
 Db 1201 TCTTCTACTATGAAAAAGGCTGATCAACGTACAGCGCGCAGCTCAG 1248
 RESULT 21
 ; US-08-504-265B-1
 ; Sequence 1, Application US/08504265B
 ; Patent No. 5837516
 ; GENERAL INFORMATION:
 ; APPLICANT: Marcus D. Ballinger and James A. Wells
 ; TITLE OF INVENTION: SUBSTITUTED VARIANTS CAPABLE OF CLEAVING
 ; TITLE OF INVENTION: SUBSTITUTED VARIANTS CONTAINING BASIC RESIDUES
 ; NUMBER OF SEQUENCES: 90
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/504,265B
 ; FILING DATE: 19-Jul-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/398028
 ; FILING DATE: 03-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kubinec, Jeffrey S.
 ; REGISTRATION NUMBER: 36,575
 ; REFERENCE/DOCKET NUMBER: P0936P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-8228
 ; TELEFAX: 650/952-9881
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 8119 base pairs
 ; TYPE: Nucleic Acid
 ; STRANDEDNESS: Single
 ; TOPOLOGY: Linear
 ; US-08-504-265B-1
 Alignment Scores:
 Pred. No.: 2,79e-21 Length: 8119
 Score: 328.50 Matches: 130
 Percent Similarity: 40.52% Conservative: 71
 Best Local Similarity: 26.21% Mismatches: 168
 Query Match: 5.85% Indels: 127
 DB: 2 Gaps: 21
 US-09-830-837-6 (1-1052) x US-08-504-265B-1 (1-8119)
 Qy 8 LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys 27
 Db 52 ATAATCTCTATTGTTATTCTGCAATGAAAAAGAGAGAGATAAGAGTGTAGAGGC 111
 Qy 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
 Db 112 AAAAAAGTATGATCATGTTGCTGTTTCTTACCG-----TTA 150

```
QY 42 ThrLeuLysValGluPheSerSerThrValVal-----52
Db 151 ATCTTTACGATGCGTTCGCCACACATCTCTGCCCGCCGCGGAAATCAACGGG 210
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 211 GAAAGAAATATATCTCGGGTTT-----AAACAGACAATGAGCAGC 252
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn 92
Db 253 ATGAGCGCGCTAAGAAGAAAGATGCAAT-----TCAGAA 288
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 289 AAAGCGCGAAAGTGAAGCAATTCAAATATGTAGAC-----GCA 330
QY 113 GlyLeuLeuThrLeuLysPheProAsnIleLysArgValThrProGlnArgLysVal 132
Db 331 GCTTCAGCTACATTAACGAA-----351
QY 133 PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
Db 352 ---AAAGCTGTAAAGAAATGAAAGACCGGAGCTGCT-----390
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySerGly 172
Db 390 -----390
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGln 192
Db 391 -----TACGTGAAGAAGATCAGCTAGCAGCATCGCTAGCGGAGTCCGCTACGGC 444
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnVal 212
Db 445 GTATCAAA---ATTAAAGCCCTGCTCTGCACTCTCAAGGCTACACTGGATCAATGTT 501
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys-----229
Db 502 AAAGTAGCGGTATCGACAGCGGTATCGATTCTCTCATCTCTTAAGGTAGCAGCG 561
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGly 246
Db 562 GGAGCCAGCATGTTCTCTTCAACAAAT-----CCTTTCCAAGACAAC 606
QY 247 LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263
Db 607 GACTCTCAGGAACTACGTTCCCGCACAGTTGCGCTCTTAATAACTCAATCGGTGTA 666
QY 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal 283
Db 667 TTAGCGGTTCGCCAAGCCATCACTTTACGCTGTAAAGTTCTCGTCTGACGGTTC 726
QY 284 SerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal 303
Db 727 GGCCAAATACAGTGTGATCAATCAAGCAATCGAGTGGCGCATGCCAAACAATATGGAGTT 786
QY 304 LeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheValAspLys 321
Db 787 ATTAACATGAGGCTCGCGGACCTCTGTTCTGCTGTTTAAAGCGGCGATGTGATAAA 846
QY 322 ValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341
Db 847 GCC-----GTTGATCCGCGCTGCTGAGTGTGCTGCTTAAAGCGGCGATGTGATAAA 897
QY 342 -----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
Db 898 TCCGCGAGCTCGTCACAGCTGAGTACCTGCTGCAATACCTCTCTCATTCAGTATGAGC 957
QY 360 GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu 379
Db 958 GCTGTTGACAGCAGCAACAAAGAGCATCTTTCTCAAGCGTAGGACCT-----GAG 1008
```

380 LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArg 399
1009 CTT-----GATGTCATGCACCTGGCGTATCTATCCAA 1041
400 GlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
1042 AGCAGCTCTCTCGAAACAAATACGGG-----CGGTACAACGGTACCTCAATGGCATCT 1095
418 ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
1096 CCGACGTTGCCGAGCGGCTGCTTGTATCTTCT-----AAGCACCGCACTGG 1146
438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn 457
1147 ACAACACACTCAAGTCCGACGAGCTTTAGAAAACACCACACATAAACTTGT-----GAT 1200
458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGln 473
1201 TCTTTCTACTATGAAAGGGCTGATCAACGTACAGCGGCGCTCAG 1248

RESULT 22
RE34606-4
; Patent No. RE34,606
; APPLICANT: ESTELL, DAVID A.; WELLS, JAMES A.; BOTT,
; RICHARD R.
; TITLE OF INVENTION: MODIFIED ENZYMES AND METHODS FOR
; MAKING SAME
; NUMBER OF SEQUENCES: 27
; CURRENT APPLICATION DATA:
; FILING DATE: 20-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 614,612
; FILING DATE: 29-MAY-1984
; SEQ ID NO: 4:
; LENGTH: 1498
; RE34606-4

Alignment Scores:
Pred. No.: 2,94e-22 Length: 1498
Score: 325.50 Matches: 99
Percent Similarity: 47.58% Conservative: 58
Best Local Similarity: 30.00% Mismatches: 128
Query Match: 5.79% Indels: 45
DB: 6 Gaps: 14

US-09-830-837-6 (1-1052) x RE34606-4 (1-1498)

QY 180 HisSerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAla 199
Db 438 CATATTGCACATGAATATGCGCAATCTGTCCTTATGGCATTTCTCAA---ATTAAAGCG 494
QY 200 AspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThr 219
Db 495 CCGGCTCTTCACTCTCAAGGCTACACAGGCTTAAGCTAAAGTAGCTGTTATCGACAGC 554
QY 220 GlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsn 239
Db 555 GGAATTTGACTTCTCTCATCTGACTTA---AAGCTCAGAGCGGCGGACGACATTCGACCT 611
QY 240 GluArgThr-----LeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIle 257
Db 612 TCTGAAACAAACCCATACCGAGCGGCGAGTCTTCACGGTACCGATCTAGCGGTACGATT 671
QY 258 AlaSerMetArgGlu-----CysGlnGlyPheAlaProAspAlaGluLeuHisIle 274
Db 672 GCGGCTCTTAATAACTCAATCGGTGTTCTGGGGTGTAGCCCAAGCGCATCATATATGCA 731
QY 275 PheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsn 294
Db 732 GTAAAGTCTGATTCAACAGAGCGGCGCAATATAGCTGGATTATTAAAGCGCATGTAG 791
QY 295 TyrAlaIleLeuLysLysIleAspValLeuAsnLeuSerIleGlyGlyPro-----Asp 312

Db	525	-----	525
Qy	171	SerGlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIlePro	190
Db	526		526
Qy	191	ArgGlnValAlaGlnThrLeuGlnAlaSerValLeuTrpGlnMetGlyTyrThrGlyAla	210
Db	559	TACGGCATT---CCCTCTCATTAAGCGGACAAAGTCGAGGCTCAAGCGCTTTAAGGAGCGC	615
Qy	211	AsnValArgValAlaValPheAspThrGlyLeuSerGlyHisProHisPheLysAsn	230
Db	616	AATGTAAAGTAGCGCTCGTGGATCAGAGAAATCCACAGCTTCTCATCCGGACTTGACACTA	675
Qy	231	ValLysGlnArgThrAsnTrpThrAsnGlnArgThrLeuAspAspGlyLeuGlyHisGly	250
Db	676	GTCGCGGAGCAAGCTTTGTGGTCGGCAAGCTTATACACCGGACGCGCAACGACACGCG	735
Qy	251	ThrPheValAlaGlyValIleAlaSerMetArgGlyCysGln-----GlyPheAla	2675
Db	736	ACACATGTTGCGGTACAGTAGCTGCCTTGACAATAACACGGGTGATTATAGCGTTGCG	795
Qy	268	ProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSer	287
Db	796	CCAACGCTACTCTTGTCGCGTTTAAGTACTGAATTCAAGCGGAGCGGATCATACACG	855
Qy	288	TrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysIleAspValLeuAsnLeuSer	307
Db	856	GGCATTGTACCGGAATCGATGGGCGACACAAACGGCATGGATGTTTCAATATGAGC	915
Qy	308	IleGlyGlyProAspPheMetAsp-----HisProPheValAspLysValTrpGluLeu	325
Db	916	CTTGGGGGAGCATCTAGGCTCGACAGCATGAAACAGGCGAGTCGACATGCATAT	969
Qy	326	ThrAlaAsnAsnValIleMetValSerAlaIleLeuGlyAsnAspGlyProLeu-----Tyr	343
Db	970	---GCCAAGAGGGGTTGTCTGTAGCTGCGACAGGAAACAGCGGATCTTCAGAAACACG	1026
Qy	344	GlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyIleAspPhe	363
Db	1027	ATATCAATTTGGCTATTCCTGCAAAATACGATTCTCATCGCTGTTGGTGGGTAGACTCT	1086
Qy	364	GluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGluLeuProGlyGly	383
Db	1087	AACAGCAACAGAGCTTCATTTTCCAGCGTCGGAGCA-----GAGCTT-----	1128
Qy	384	TyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGlySerGlyVal	403
Db	1129	-----GAAGTCAGTGGCTCTGCGCAGGCGTATACAGCACTTACACCA	1170
Qy	404	LysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValAlaGlyAla	423
Db	1171	ACGAACACTTATCCACACTTACAGGAAAGCTCATGGCTTCTCTCTATGTAGCGGAGCA	1230
Qy	424	ValThrLeuLeuValSer---ThrValGlnLysArgGluLeuValAsnProAlaSerMetLys	443
Db	1231	CGACGTTTGAATCTGTCAAAATCCGAACCTTCACGCTT-----CACAA	1275
Qy	443	sGlnAlaLeuIleAlaSerAlaArgArgMetProGlyValAsnMetPheGluGlnGlyHI	463
Db	1276	GTCCCAACAGCTCTCTCCAGCAGCGGCACTTATTGGGA-AGCTCTCTTACTATGGGAA	1334
Qy	463	sGlyLysLeuAspLeuArgAlaTyrGln	473
Db	1335	AGGTCGTATCAATGTGGAAGCTGCGCTCCATCA	1365

RESULT 25

US-09-000-016-3

; Sequence 3, Application US/09000016

; Patent No. 6143541

; GENERAL INFORMATION:

; APPLICANT: AKIRA ARISAWA et al.

```
QY 188 ---AlaIleProArgGlnValAlaGlnThr-LeuGlnAlaAspValLeuTrpGlnMetG1 206
Db 940 CAGGGCGCGCTCGACACGCTCCGTCGGGCGACAGTCCGCCAAGGCGTGGTCCGGCGG 999
QY 206 yTyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGlyLysHisPr 226
Db 1000 CTACGACGGCAAGGGGTGAAGTCCGCTCGGACACCGGTGTGACACAGCCATCC 1059
QY 226 oHisPheLys---AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAs 245
Db 1060 GGACCTCAAGGGCGGGGTGACCGCTCCAAGAACTTCACGCGCGCGCGCGCGCGCGA 1119
QY 245 pGlyLeuGlyHisGlyThrPheValAlaGlyValIleAla-----SerMe 260
Db 1120 CAAGGTGGGCGACGGCACCCACGCTCGCTCGGGCGGCGACGGCGCGCGCGCGCA 1179
QY 260 tArgGluCysGlnGlyPheAlaProaspAlaGluLeuHisIlePheArgValPheThrAs 280
Db 1180 GGCAAGTACAGGGGTGTCACCGCGCGCGGATCCTCAAGCGCAAGGTCTCTCGACGA 1239
QY 280 nAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLy 300
Db 1240 CTCGGTTCGGCGACGACTCCGGCATCTCGCGCGCATGGAGTGGCGCGCGCGCAGGG 1299
QY 300 sIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProPheValas 320
Db 1300 CGCGGACGTCTCACCATGAGCTGGGGCGC-----ATGGACACACCGGAGACCGA 1350
QY 320 p-----LysValTrpGluLeuThrAla---AsnAsnValIleMetValSerAl 335
Db 1351 CCGGTGGAGCGCGGTGCACAAAGCTGTCGCGCGAGAGGGCGTCTGTTCGCCATCGC 1410
QY 335 aIleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspVa 355
Db 1411 GGCGGCAAGAGGGCGC-----GAGTCGATCGGTTCGCCCGGCGCGCGCGCGCGC 1464
QY 355 lIleGlyValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMe 375
Db 1465 CTTACCGTGGCGCGGTGACGACACAGGACAAAGTCCGCGACTTCTCTCCACCGGCC 1524
QY 375 tThrThrTrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrG1 395
Db 1525 CGGCCTC-----GGCGACGGCGCCATCAAGCGGAGCTCACCCTCCCGC 1569
QY 395 yAlaGlyValArgGlySerGlyValLysGly----- 405
Db 1570 CFTGGACATCAGCGCGCTCGCGGAGGGCAAGCATCGGCCAGGAGTCTGGTGAAGG 1629
QY 406 -----GlyCysArgAlaLeuSerGlyThrSerValAlaSerProValValAlaGlyAl 423
Db 1630 ACCGGCGGCTACATGACCATCTCCGCGACGTGATGGCGACCGCGCGCGCGCGC 1689
QY 423 aValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSerMetLy 443
Db 1690 GGCGGCGCTCTG-----AAGCAGCAGACCCCGGACTGGACCTCCGCGCAACTGAA 1740
QY 443 sGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAlaAsnMetPheGluGlnGlyH1 463
Db 1741 GGCGCGCTCACCAGTCCACCAAGGCG---GGCAAGTACACCCCGCTTCAGCAGGGTTC 1797
QY 463 sGlyLysLeuAspLeuArgAlaTyrGln 473
Db 1798 GGCGCGGATCCAGGCGGACAAAGCGCTCCAG 1828
```

Search completed: June 2, 2003, 05:46:17
Job time : 337 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 2, 2003, 01:15:11 ; Search time 535 Seconds
(without alignments)
4428.228 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNIVLLVLLCGKKH.....PRVKPQLMQVHPKPTSPV 1052

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USFTO.spool/US09830837/runat_23052003_181923_8059/app_query.fasta_1.1223
-DB=N.Geneseq_101002 -QFIX=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=25
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830837 @CEN_1_1_344 @runat_23052003_181923_8059 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5617	100.0	4338	21	Human subtilisin-k
2	5617	100.0	4338	24	Human site-1 prote
3	5487	97.7	4198	24	Hamster site-1 pro
4	5476	97.5	3895	21	Rat subtilisin-kex
5	5427	96.6	3788	21	Mouse subtilisin-k
6	3560	63.4	2729	24	Human polynucleoti
7	2366.5	42.1	3160	22	Drosophila site 1
8	2231	39.7	2859	23	Drosophila melanog
9	2219.5	39.5	5332	23	Gastric cancer ass
10	1229.5	21.9	786	20	Human gene express
11	1155	20.6	757	20	Human polynucleoti
12	474.5	8.4	469	22	Human excretory re
13	348	6.2	9689	22	Human kidney relat
14	348	6.2	9689	22	Coding region of t
15	341	6.1	1220	9	apr gene encoding
16	341	6.1	1220	11	Sequence of the ap
17	341	6.1	1524	7	DNA encoding the a
18	339	6.0	2044	9	Subtilisin gene fr
19	338.5	6.0	1500	8	B.subtilis subtili
20	338.5	6.0	1500	16	Bacillus subtilis
21	338.5	6.0	1500	20	Subtilisin gene.
22	338	6.0	1499	11	Bacillus subtilis
23	338	6.0	2760	20	Bacillus subtilis
24	337.5	6.0	1180	22	Subtilisin E.DNA,
25	337.5	6.0	1500	8	Bacillus subtilis
26	336	6.0	266	20	Human gene express
27	335.5	6.0	1140	24	Bacillus lichenifo
28	334	5.9	1220	8	Thermotable Bacil
29	332.5	5.9	1146	24	Bacillus subtilis
30	331.5	5.9	1494	24	Bacillus amyloliqu
31	331	5.9	1497	18	Bacillus amyloliqu
32	331	5.9	1497	18	Bacillus amyloliqu
33	331	5.9	1497	20	Bacillus amyloliqu
34	331	5.9	1497	20	B. amyloliquefacie
35	331	5.9	1497	20	B. amyloliquefacie
36	331	5.9	1497	20	Subtilisin DNA seq
37	331	5.9	1497	21	Subtilisin DNA seq
38	331	5.9	1497	21	Bacillus amyloliqu
39	331	5.9	1497	24	Bacillus amyloliqu
40	330	5.9	1496	8	Streptomyces virid
41	330	5.9	2539	18	Dhpa-mel chimeric
42	330	5.9	2809	18	B. amyloliquefacie
43	329.5	5.9	1492	16	B. amyloliquefacie
44	328.5	5.8	1494	20	Subtilisin gene.
45	328.5	5.8	1494	20	B. amyloliquefacie

ALIGNMENTS

RESULT 1
AAA57197
ID AAA57197 standard; cDNA; 4338 BP.
AC AAA57197;
XX
XX
XX
DT 03-OCT-2000 (first entry)
XX
DE Human subtilisin-kexin isoenzyme 1 cDNA.
XX
KW Human; subtilisin-kexin isoenzyme 1; SKI-1; antilipase; cytostatic;
KW vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
KW Ras-dependent cancer; restenosis; amyloid protein formation;
KW pro-brain-derived neurotrophic factor; proBDNF;
KW steroid-regulatory element-binding protein; SREBP; ds.
OS Homo sapiens.
XX

QY	461	GlnGlyHisGlyLysLeuAspLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro	480
Db	1877	CAAGGCCAGCGCAAGCTCGCTCAGAGCCTATCAGATCCTCAACAGCTACAGCCA	1936
QY	481	GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr	500
Db	1937	CAGGCAAGTTTGAGCCCAAGCTACATAGATCTGACTAGTGTCCCTACATGTGGCCCTAC	1996
QY	501	CysSerGlnProIleTyrGlyMetProThrValValAsnValThrIleLeuAsn	520
Db	1997	TGCTCCAGCCCATCTACTATGAGGAATGCCAGAGTTGTAATCTCACCATCCTCAAC	2056
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln	540
Db	2057	GGCATGGAGTCACAGGAAGATTGTAGATAAGCTGACTGACGAGCCCTATTGGCCACAG	2116
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGly	560
Db	2117	AACGGAGACAACATTGAAGTTGCCTCTCTACTCTCGGTCTTATGGCCTTGGTCGGGC	2176
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln	580
Db	2177	TACCTGGCCATCTCCATTTCTGTGACCAAGAAGCGCTTCTCTGGGAAGCATTTGCTCAG	2236
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600
Db	2237	GGCCATGTCATGATCCTGCTTCCCCAGCAGACAGAGTCACAAAATGGTGCAGAA	2296
QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleLeuProThrProArgSer	620
Db	2297	CAGACTCAACAGTAAGCTCCCATTAAGGTGAAGATAATTCCTACTCCCGCGCAAGC	2356
QY	621	LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArg	640
Db	2357	AGAGAGTTCTCTGGGATCAGTACCAACCTCCGCTATCCACTGCTTATTCGCCAGG	2416
QY	641	AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660
Db	2417	GATAATTTAAGATGAAGAATGACCTTTAGACTGGAATGGTGCATCCACACCAAT	2476
QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
Db	2477	TTGAGGGATATGACAGCATCTGAGAAGCATGGGCTACTTTGTAGAGTCTCTGGGGCC	2536
QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
Db	2537	CCCTTCACGTGTTTGTATGCCAGTCAGTATGACCTTCTGATGGTGACAGTGAGGAG	2596
QY	701	GluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu	720
Db	2597	GAGTACTTCCCTGAAGAGATCGCAAGCTCCGGAGGAGCTGGACAACGGCTCTCGCTC	2656
QY	721	ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
Db	2657	GTGATCTTCAGTGACTGGTACAACTCTCTGATGAGAAAAGTGAAGTTTATGATGAA	2716
QY	741	AsnThrArgGlnTrpTrpMetProAspThrGlyValAlaAsnIleProAlaLeuAsnGlu	760
Db	2717	ACACACAGCAGTGTGGATGCGGATACCGGAGGAGCTACATCCAGCTCTGAAATGAG	2776
QY	761	LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu	780
Db	2777	CTGCTCTCTGTGGAAACATGGGTTACGCGATGGCTGTATGAAGGGAGTTCACCTG	2836
QY	781	AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800
Db	2837	GCCAAACCATGACATGATTTATGCTCAGGGTCAGCATCGCAAGTTTCCAGAAGATGGC	2896
QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValIleLysGlnGluThrAla	820
Db	2897	GTGCTGATACACAGACTTTCAGGACCAAGGATTGGAGGTTTAAAGCAGGAACAGCA	2956
QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
Db	2957	GTGTTTGAACAGCTCCCATTTTGGGACTTTATCAGATTCACAGTTCAGGCTGAGGCGG	3016
QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
Db	3017	ATTGTACTATGGGGACTCCAATTGCTTGGATGACAGTCACCGACAGAGGACTGCTTT	3076
QY	861	TtpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer	880
Db	3077	TGCTTCTCGATGCGCTCTCTCCAGTACATCGTATGGGTGACACCGCTAGCTCAGT	3136
QY	881	HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet	900
Db	3137	CACCTCTGGGAACCGCCAGCGCCCTCCAGTGGAGCAGGCTCAGTCACTCCAGAGAGATG	3196
QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
Db	3197	GAAGGAACCATCTTCCATCGTACTCCAAGTTCTGGAGGCCATTTGGAGAGCCCAAAA	3256
QY	921	ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu	940
Db	3257	CCTCGGCCCTTACCAGCCCTGTCACGCTTGTCTTGGGCCCAAGCCAGCCCTTAAACGAG	3316
QY	941	ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuSerIleAspLeuAspLysVal	960
Db	3317	ACGGGCCCCAGTAACCTTTGGAAACATCAGAAGCTACTCTCCATTGACCTGGACAAGTG	3376
QY	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer	980
Db	3377	GTGTTTACCACAACTTTCGATCGAATCGCCTCAAGTAGGGCCCTTGTCCCTTGAGAGAGC	3436
QY	981	GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln	1000
Db	3437	GGCGCTGGACATCTCTGAGGGATCATGCTGGCCGCTACAAACAGGAGGTGGCCAG	3496
QY	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValGln	1020
Db	3497	ACCATTCCTGCTCTTTCCTTCCCTGGAGCCATGGTGTCTCTGGCTTCTTTGTGTACAA	3556
QY	1021	IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu	1040
Db	3557	ATCAACAGGCCCAAGAGCAGGCCGAAGCGGAGGAGGCCAGGCTGAAGCGCCGCGCAGCTC	3616
QY	1041	MetGlnGlnValHisProProLysThrProSerVal	1052
Db	3617	ATGCACAGGTTCCACCGCCCAAGACCCCTTCGGTG	3652
RESULT 2			
AAD24182			
ID	AAD24182 standard; DNA; 4338 BP.		
XX			
AC	AAD24182;		
XX			
DT	07-MAY-2002 (first entry)		
XX			
DE	Human site-1 protease DNA.		
XX			
KW	Site-1 protease; SLP; promoter; transcription; obesity; type II diabetes;		
KW	sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;		
KW	dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.		
XX			
OS	Homo sapiens.		
XX			
PH	Location/Qualifiers		
FT	497..3655		
FT	/*tag= a		
FT	/product= "Site-1 protease"		
XX			
PN	WO200200873-A1.		
XX			
PD	03-JAN-2002.		
XX			

PF 19-JUN-2001; 2001WO-SE01386.
 XX
 PR 27-JUN-2000; 2000SE-0002417.
 XX
 PA (BIOV-) BIOVITRUM AB.
 XX
 PI Abrahamson L, Ekblom J, Forsgren M, Hoerling J, Johansson P;
 DR WPI; 2002-139918/18.
 DR P-PSDB; AAEL14527.
 XX
 XX Human site-1 protease promoter region for identifying agents capable of
 PT inhibiting the promoter activity useful in treating medical conditions
 PT such as obesity, diabetes, atherosclerosis and hypercholesterolemia -
 XX
 XX Example 1; Page 16-21; 36pp; English.
 PS
 CC The invention relates to human site-1 protease (SIP) promoter region.
 CC The promoter sequence is useful for identification of compounds that
 CC inhibit transcription of SIP, which in turn results in inhibition
 CC of steroid regulatory element-binding protein (SREBP) pathway. The
 CC compound identified is useful for the treatment of medical conditions
 CC related to obesity, type II diabetes, hypercholesterolemia,
 CC dyslipidemia, atherosclerosis and other cardiovascular diseases.
 CC The present sequence is human SIP DNA.
 XX
 SQ Sequence 4338 BP; 1085 A; 1051 C; 1160 G; 1042 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 4338
 Score: 5617.00 Matches: 1052
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0
 US-09-830-837-6 (1-1052) x AAD24182 (1-4338)
 QY 1 MetLysLeuValAsnIleTrpLeuLeuValValLeuLeuCysGlyLysLysHis 20
 DB 497 ATGAAGCTTGTCAACATCTGGCTGCTCTGCTGCTGCTGCTGGGAGAACAT 556
 QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40
 DB 557 CTGGCGCAGACTGGAAAGAAATCTTTGAAAGGCCCATGCCCTGGCTGTCCTCCAC 616
 QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrIleValAlaPhe 60
 DB 617 CTGACTTTGAAGTGGGAATCTCATCAACAGTGTGGAAATGAATATATATGPGCTTTC 676
 QY 61 AsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSer 80
 DB 677 AATGGATACTTTACAGCCAAAGCTAGAAATTCATTTATTTCAAGTGCCTCGAAGACAGT 736
 QY 81 GluValAspAsnTrpArgIleIleProArgAsnProSerSerAspTyrProSerAsp 100
 DB 737 GAAGTAGAATTTGGAGAAATATACCTCGAAACAAATCCATCCAGTACTACCTAGTCAT 796
 QY 101 PheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis 120
 DB 797 TTTGAGGTGATTTCAGATAAAAGAAAAACAGAAAGCGGGCTGCTAACACTTGAAGATCAT 856
 QY 121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu 140
 DB 857 CCACACATCAACGGGTACAGCCCAACGAAAGAGTCTTCGTTCCCTCAAGTATGCTGNA 916
 QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg 160
 DB 917 TCTGACCCCAAGTACCTTCGCAATGAACCCGGTGGAGCCAGAGGTCATCATCAGT 976
 QY 161 ProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis 180
 DB 977 CCCCTGCGAAGACCCAGCCTCTCCCTGGGCTCTGCGCTTCTGGCATGCTACGGGAAGGCAT 1036

QY 181 SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
 DB 1037 TCGAGCAGACGGCTGCTGAGAGCCATCCCGCGCCAGGTTCGCCAGACACTGCAGCGAGAT 1096
 QY 201 ValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGly 220
 DB 1097 GTGCTCTGCCAGATGGGATATACAGTGTCTAATGAAGAGTTCCTTTTGTGACATGGG 1156
 QY 221 LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240
 DB 1157 CTGACCGAGAGAGCATCCACCTTCAAAATCTGAAGCAGAGAACCACTGGACCAAGAG 1216
 QY 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
 DB 1217 CGAAGCTGGACAGTGGTGGCCATGGCACATCTCGGCAGTGTGTATAGCCAGCATG 1276
 QY 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
 DB 1277 AGGGAGTCCCAAGGATTTGCTCAGATCGAAGCTTCACATTTTCAGGCTCTTTACCAAT 1336
 QY 281 AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLys 300
 DB 1337 AATCAGGTATCTTACACATCTGGTGTGGACGCCCTCAACTATGCCATTTTAAAGAG 1396
 QY 301 IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisProPheValAsp 320
 DB 1397 ATCGACGTGTAAACCTTCAGCATCGCGCGCCGACCTCATGTGATCATCCGTTTGTGAC 1456
 QY 321 LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly 340
 DB 1457 AAGGTGTGGGAATTAACAGCTAACCAATTAATCATGTGTTCTGTCTATTTGGAATCAGGA 1516
 QY 341 ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 360
 DB 1517 CCTCTTTATGGCACTCTGAATAACCTCTCATCAATGGATGTGATGGTAGTAGCGGC 1576
 QY 361 IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGluLeu 380
 DB 1577 ATTGACTTGAAGATAACATCCCGCGCTTTCTTCAAGGGGAATGACTACCTGGAGCTA 1636
 QY 381 ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly 400
 DB 1637 CCAGGAGCTACGGTCCATGAACCTGACATGTCTACCTATGTGCTGGCTGGCGGGT 1696
 QY 401 SerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420
 DB 1697 TCTGGCGTGAAGGGGGGTGCCGGGCCCTCTCAGGGACCAAGTGTGCTTCTCCAGTGTT 1756
 QY 421 AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAla 440
 DB 1757 GCAGGTGCTGTCACTTGTGTAGTGACACAGTCCAGAAAGCGTGAGTGGTGAATCCCGCC 1816
 QY 441 SerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMetPheGlu 460
 DB 1817 AGTATGAAGCAGGCGCTGATCGCTCAGCGGAGGTCCTCCGGGGTCAACATGTTTGAG 1876
 QY 461 GlnGlyHisGlyLysLeuAspLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480
 DB 1877 CAAGGCCACGGCAAGCTCGATCTGCTCAGAGCCCTATCAGATCTCAACAGCTACAGCCA 1936
 QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
 DB 1937 CAGGCAAGTTTGAGCCCGCCAGCTACATAGTCTGACTGAGTGTCCCTACATGTGGCCCTAC 1996
 QY 501 CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn 520
 DB 1997 TGCTCCCGCCCATCTACTATGGAGAAATGCCAGAGTGTATATGTCACTCTCTCAAC 2056
 QY 521 GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln 540
 DB 2057 GGCATGGGAGTCAAGAGAAATTTAGATAAGCCTGACTGGAGCCCTATTTTCCACAG 2116

QY 541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTyrSerGly 560
DB 2117 AACGGAGACAACATTGAAGTTGCTTCTCTACTCTCTGCTCTATATGCGCTTGGTGGGC 2176
QY 561 TyrIleuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln 580
DB 2177 TACCTGGCCATCTCCATTCTGTGACCAAGAAGACGGCTTCTTGGGAAGCATTTGCTCAG 2236
QY 581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyValAlaGlu 600
DB 2237 GGCCATGTCATGATCACTGTGGCTTCCCGAGCAGACAGATCAAAAATGTTGTCAGAA 2296
QY 601 GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProArgSer 620
DB 2297 CAGACTTCAACAGTAAAGCTCCCAATTAAGGTGAAGATAATCTACTCCCGCGGAAGC 2356
QY 621 LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArg 640
DB 2357 AAGAGAGTCTCTGGGATCAGTACCACAACCTCCGCTATCCACTGGCTATTTCCCGCAGG 2416
QY 641 AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn 660
DB 2417 GATAATTAAAGGATGAAGATGACCTTTAGACTGGAATGGTATCACATCCACACCAAT 2476
QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
DB 2477 TTCAGGGATATGTACCAGCATCTGAGAAGCATGGCTACTTTGTAGAGGTCTCTCGGGGCC 2536
QY 681 ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
DB 2537 CCCTTCACGTGTTTGAATGACATGATGGCACTTTGCTGATGGTGGACAGTAGGAG 2596
QY 701 GluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu 720
DB 2597 GAGTACTCTCCTAGAGATGCGCAAGCTCCGGAGGACGTGCACACAGCCCTCTCGCTC 2656
QY 721 ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
DB 2657 GTCATCTTCAGTGACTGGTACAACTCTCTGTATGAGAAAGTGAAGTTTATGATGAA 2716
QY 741 AsnThrArgGlnTrpTrpMetProAspThrGlyValAlaAsnIleProAlaLeuAsnGlu 760
DB 2717 AACACAGGCGATGGTGGATGCCGATACCGAGGAGCTAACATCCAGCTCTGGAATGAG 2776
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
DB 2777 CTGCTGCTGTGTGGAACATGGGGTTCAGCGATGGCTGTATGAAGGGAGTTACCCCTG 2836
QY 781 AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
DB 2837 GCCAACCATGACATGATATGCTGCTCAGGGTGCAGCATCCGAAAGTTTCCAGAAGATGC 2896
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnThrAla 820
DB 2897 GTCTGTATACACAGACTTTCAGGACCAAGGATGGAGTTTAAAGCAGGAACAGCA 2956
QY 821 ValValIleAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
DB 2957 GTTGTGTAACACGCTCCCATTTTGGGACTTATCAGATTCACACTGAGGGTGGAGCCGG 3016
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
DB 3017 ATTGTACTGTATGGGACTCCAATTGCTTGGATGACAGTCAACGACAGAAGGACTGCTT 3076
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
DB 3077 TGGCTTCTGGATGCCCTCTCCAGTACATCTATGGGTGACACCGCTACCCCTCAGT 3136
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
DB 3137 CACTCTGGNACCCGAGGCCCTCCAGTGGAGCAGGCTCAGTCACTCCAGAGAGATG 3196
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920

DB 3197 GAAGGAAACCATCTTCTATCGTACTCCAAAGTTCTGGAGGCCATTTGGGAGACCCAAA 3256
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
DB 3257 CCTCGGCTCTACAGCCTCTCACGCTTCTTGGGCCAAGCCACAGCCTTTAAACGAG 3316
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
DB 3317 ACGGCGCCCAAGTAACTTTGGAAACATCAGAAGCTACTCTCCATTGACCTGGACAAAGTG 3376
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
DB 3377 GTGTTACCCCAACTTCGATCGAATCGCCCTCAAGTAGAGCCCTTGTCCCTGGAGAGAGC 3436
QY 981 GlyValTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
DB 3437 GCGCCTGGGACATCTCTGGAGGATCATCCCTGGCCGCTACAAACAGGAGGTGGGCCAG 3496
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
DB 3497 ACCATTCTCTCTTGGCTTCTCTGGAGCCATGGTGTCTCTGGCTTCTTTGGGTACAA 3556
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgLysArgProGlnLeu 1040
DB 3557 ATCAACAAGGCCAAGAGAGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3616
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
DB 3617 ATCAGCAGGTTCCCGCCCAAGACCCCTTCGGTG 3652
RESULT 3
AAD24185
ID AAD24185 standard; DNA; 4198 BP.
XX
AC AAD24185;
XX 07-MAY-2002 (first entry)
XX
XX Hamster site-1 protease DNA.
DE
XX Site-1 protease; SLP; promoter; transcription; obesity; type II diabetes;
KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;
KW dyslipidaemia; atherosclerosis; cardiovascular disease; hamster; ds.
XX
OS Cricetulus griseus.
XX
FH Key Location/Qualifiers
FT CDS 387..3545
FT /*tag= a
FT /product= "Site-1 protease"
XX
XX W0200200873-A1.
XX
XX 03-JAN-2002.
XX
XX 19-JUN-2001; 2001WO-SE01386.
XX
XX 27-JUN-2000; 2000SE-0002417.
XX
XX (BIOV-) BIOVITRUM AB.
XX
XX Abrahamsen L, Ekblom J, Forsgren M, Hoerling J, Johansson P;
DR WPI; 2002-139918/18.
DR P-PSDB; AAE14528.
XX
XX Human site-1 protease promoter region for identifying agents capable of
PT inhibiting the promoter activity useful in treating medical conditions
PT such as obesity, diabetes, atherosclerosis and hypercholesterolemia -
XX
PS Disclosure; Page 24-29; 36pp; English.
XX

CC The invention relates to human site-1 protease (SLP) promoter region.
 CC The promoter sequence is useful for identification of compounds that
 CC inhibit transcription of SLP, which in turn results in inhibition
 CC of sterol regulatory element-binding protein (SREBP) pathway. The
 CC compound identified is useful for the treatment of medical conditions
 CC related to obesity, type II diabetes, hypercholesterolemia,
 CC dyslipidemia, atherosclerosis and other cardiovascular diseases.
 CC The present sequence is hamster SLP DNA.

XX Sequence 4198 BP; 1061 A; 1014 C; 1106 G; 1017 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 4198
 Score: 5487.00 Matches: 1021
 Percent Similarity: 98.57% Conservative: 16
 Best Local Similarity: 97.05% Mismatches: 15
 Query Match: 97.69% Indels: 0
 DB: 24 Gaps: 0

US-09-830-837-6 (1-1052) x AND24185 (1-4198)

QY 1 MetLysLeuValAsnLlePLeuLeuLeuValValLeuLeuCysGlyLysLysHis 20
 DB 387 ATGAAGCTCATCAACATCTGGCTTCTCTGCTGGTGTTCCTCTGGGAAGAAGCAT 446
 QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40
 DB 447 CTGGGTGACAGGCTGGGGAAGACGGTTGAAGAGCATATGCCCTAGCTGTCCCCAC 506
 QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrIleValAlaPhe 60
 DB 507 CTGACTTTGAAGTGGGAATCTCCCTCAACTGTGGTGAATATGAATATGTGGCTTTC 566
 QY 61 AsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSer 80
 DB 567 AACGGATATCTCAGCCAAAGCTAGAACTCATATTTTCAAGTGTCTGCAAAAGCAGT 626
 QY 81 GluValAspAsnTrpArgIleIleProArgAsnProSerSerAspTyrProSerAsp 100
 DB 627 GAAGTAGAACACTCGAATATTATACCTCGGAACAACCCATCCAGTACTACCTAGTAGT 686
 QY 101 PheGluValIleGlnIleLysGluLysGlnLysAlaGluLeuThrLeuGluAspHis 120
 DB 687 TTTGAGGTGATTACAGATAAAGAGAAGCAGAGCGCGGCTGCTCACACTTGAAGATCAT 746
 QY 121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu 140
 DB 747 CCMAACATCAAGCGGTGACACCTCAGCCAAAGTCTTCTGTTCTTGAGTGTTCGTA 806
 QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg 160
 DB 807 TCTGACCCCAFTGTGCCATGTATGAACACTCGGTGGAGCCAGAGTGGCAGTCAACGA 866
 QY 161 ProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis 180
 DB 867 CCCCTGAGAAGAGCCAGTCTCTCCCTGGGCTCTGGATTCTGGCATGTCAACAGGAAGACAT 926
 QY 181 SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
 DB 927 TCAAGCGCGGATTTGTTGAGAGCCATTCTCGACAGGTTGCCAGACATTTGCGAGCAGAT 986
 QY 201 ValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGly 220
 DB 987 GTGCTGTGGCAGATGGATACACAGTGTCTAATGTCCAGGGTTGCTGTTTGTATACCTGG 1046
 QY 221 LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240
 DB 1047 CTCAGTGAGAGCATCCACACTTCAGANATGTGAAGAGAGAACAACACTGGACCAATGAG 1106
 QY 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
 DB 1107 CGACCCCTGGATGTGGCTGGCCATGGCACATTTGTCCAGGTGTGATGGCCAGCATG 1166

QY 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
 DB 1167 AGGGAGTCCAGGATTTGCCAGATGCCAGATGCACATCTTCCGGGTCTTTTACCAC 1226
 QY 281 AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLys 300
 DB 1227 AATCAGGTGTCTTACACATCTTGGTGTGGACGCTTCACTATGCCATCTCAAGAG 1286
 QY 301 IleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProPheValAsp 320
 DB 1287 ATTGATGTTCTAAACCTTAGCATCGCGGGCTGACTTCATGATCATCCCTTTTGTGAC 1346
 QY 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
 DB 1347 AAGGTCTGGGAATTAACAGCTAACATGTATCATGTTCTCTCTATCGGCATCATGGA 1406
 QY 341 ProLeuTyrGlyThrLeuAsnProAlaAspGlnMetAspValIleGlyValGlyGly 360
 DB 1407 CCTCTTATGGCACTCTGAATACCCAGCTGATCAGATGGATGTGATGGATGGGTGGC 1466
 QY 361 IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThrTrpGluLeu 380
 DB 1467 ATTGACTTTGAAGATAACATCGCCGCTTCTTCCAGGGGAATGACTACCTGGGAAC 1526
 QY 381 ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly 400
 DB 1527 CCAGAGGCTATGTCGCGTGAACCTGACATGTGTACCTATGTGTCGGAGTGGGGT 1586
 QY 401 SerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420
 DB 1587 TCCGGTGTGAAGGGGCTGCCGGGCACTCTCAGGGACCACTGCTCTCCACATGGT 1646
 QY 421 AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAla 440
 DB 1647 GCTGGGCTGTCACTTGTGTAGTAAAGCAGTGCAGAGCGGGAGCTAGTGAATCTCTGCC 1706
 QY 441 SerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGlu 460
 DB 1707 AGTGTGAAGCAAGCCCTGATTGTCATCAGCCGGAGGCTCTCTGCTGTAACTATGTCAG 1766
 QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480
 DB 1767 CAAGGCCATGCCAAGCTGATCTGTCGAGCCTATCAGATCTCTCAGCAGCTACAAACCA 1826
 QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
 DB 1827 CAGGCGACTTGACTCTAGTACATCAGCTGACTGAGTGCCTGCTACATGTGGCTTAC 1886
 QY 501 CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn 520
 DB 1887 TGTCTCAGCCATCTACTATGGAGGAATGCCAAATGTATGTATCTCACCATCTCAAT 1946
 QY 521 GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln 540
 DB 1947 GGCATGGGAGTTCACAGGAAGAAATTTGTGGATAAGCCTGAGTGGCGGCCCTATTATAC 2006
 QY 541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuLeuTrpProTyrSerGly 560
 DB 2007 AATGAGACAACATTTGAAGTGGCTCTCTACTCTCTCAGTGTATGGCTTGTGGTCA 2066
 QY 561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln 580
 DB 2067 TACCTGGCATCTCCATTTCTGTGACCAAGAAGCAGCTTCTCTGGAAGCATTCACAG 2126
 QY 581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu 600
 DB 2127 GGTACATCATGATCATCGGTGGCTTCCCCAGCAGAGCAGCAAGCAAAAATGTTGTCGAG 2186
 QY 601 GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProArgSer 620
 DB 2187 CATATCTCCAGTGAAGCTTCCATTAAGGTGAAGATCATTCACACCCCTCTCCGAGC 2246
 QY 621 LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArg 640

Db 2247 AAGAGAGTCTCTGGGACCATGATCACAACTCCGCTACCCCGACGGCTACTTTCCAGG 2306
QY 641 AspAsnLeuArgMetLysAsnAspProLeuAspTyrAsnGlyAspHisIleHisThrAsn 660
Db 2307 GACAACCTGGGATGAGATGATCTTTAGACTGGAATGGCCACCATGTCACACCAAT 2366
QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
Db 2367 TTCAGGACATGTACCAGCACCTGCGCAGCATGGCTACTTCTGTGAGGTGCTCGGTGCC 2426
QY 681 PropheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
Db 2427 CCATTTCACGTGCTTGTATGATACACAGTATGGCACTTGTCTATGGTGGATGTAAGAA 2486
QY 701 GluTyrPheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu 720
Db 2487 GAGTACTTCCAGGAGATGTCAGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2546
QY 721 ValIlePheSerAspTyrPyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
Db 2547 GTCATCTTCAGTCAGTACACACTTCTGTTATGAGAAAGTGAAGTTTACGATGAA 2606
QY 741 AsnThrArgGlnTyrPyrMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu 760
Db 2607 AACACAAGGCGAGTGGTGGATGCGAGATCTGGAGGAGCCAAACATCCAGCTCTGAACGAG 2666
QY 761 LeuLeuSerValTyrAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
Db 2667 CTGCTGCTGTGTGGAACTGGGTTTCAGCGATGGCTTATGAGGGGAGTTGGCCCTG 2726
QY 781 AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
Db 2727 GCGAATCATGACATGATTATGATCATCGGATGCGAGTCCCAAGTTTCCAGAAGATGGT 2786
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
Db 2787 GTTGTGATCACACAGACTTTCAGGACCAAGGATGGAGTCTTAAACAAGAGACGCA 2846
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
Db 2847 GTTGTGAAATGTTCCATTTTGGGGCTTATACAGATCCAGCTGAAGTGGGGCCGG 2906
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
Db 2907 ATCCTGTTGTATGGATTCATTCATGATGATGATGATGATGATGATGATGATGATG 2966
QY 861 TrpLeuLeuAspAlaLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
Db 2967 TGGCTTCTGGATGATCTCTTACATACATCATATATGGCTGAACCTCCAGCCTCAGC 3026
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
Db 3027 CATTCAGGACCGGAGCGCCACCGAGTGGAGTGGCTGGCCCTCTCGAAGAGATG 3086
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
Db 3087 GAAGGAAACACCTTCATGATACCTCAAGTCTTGGAGCCATCTGGAGACCCAAAA 3146
QY 921 ProArgProLeuProAlaCysProArgLeuSerThrAlaLysProGlnProLeuAsnGlu 940
Db 3147 CTTGCGGCTCTTCCAGCCTGTCCACACTTGTATGGGCCCAAGCCACAGCCTTTGAATGAG 3206
QY 941 ThrAlaProSerAsnLeuThrLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
Db 3207 ACTGCCCCAGTAATCTTGGAAACATCAGAGCTGCTCTCCATTACCTGGGAAAGTA 3266
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3267 GTGTTACCAACTTTCGATCGAATGCCCTCAAGTGAGACCTTGTCCCTCGGAGAAAGT 3326
QY 981 GlyAlaThrAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000

Db 3327 GGTGCTGGGACATCTCTGGAGGATCATGCTCGTGGCGCTACAAACCAAGAGGTGGCCAG 3386
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
Db 3387 ACCATCCCTGCTCTTGGCTTCTTCGGAGCCATGGTGGCCCTGGCCTTCTTTGGGTACAG 3446
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3447 ATCAGCAAGGCCAAAGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3506
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3507 ACACAGCAGACCCACCACCCAGGACCCCGCTCAGTG 3542

RESULT 4

AAA57198
ID AAA57198 standard; cDNA; 3895 BP.
XX
AC AAA57198;
XX
DT 03-OCT-2000 (first entry)
XX
DE Rat subtilisin-kexin isoenzyme 1 cDNA.
XX
KW Rat; subtilisin-kexin isoenzyme 1; SKI-1; antilipemic; cytostatic;
KW vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
KW Ras-dependent cancer; restenosis; amyloid protein formation;
KW pro-brain-derived neurotrophic factor; proBDNF;
KW sterol-regulatory element-binding protein; SREBP; ds.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT CDS 418..3576
FT /*tag= a
FT /product= "SKI-1"
XX
PN WC200026348-A2.
XX
PD 11-MAY-2000.
XX
PF 04-NOV-1999; 99WO-CA01058.
XX
PR 04-NOV-1998; 98CA-2249648.
XX
PA (RECL-) INST RECH CLINIQUES MONTREAL.
XX
PI Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
XX
DR WPI; 2000-365601/31.
DR P-PSDB; AAB06335.
XX
PT Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
PT producing a polypeptide useful for treating hypercholesterolemia, liver
PT steatosis and amyloidosis, comprises a specific amino acid sequence -
PS
PS Example 1; Page 61-66; 119pp; English.
XX

CC The present sequence encodes rat subtilisin-kexin isoenzyme 1 (SKI-1),
CC a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
CC from rat adrenal glands by RT-PCR using active site degenerate
CC primers. SKI-1 cleaves at a specific threonine residue within
CC the N-terminal segment of pro-brain-derived neurotrophic factor
CC (proBDNF). It is also capable of cleaving sterol-regulatory
CC element-binding proteins (SREBPs), which function to control lipid
CC biosynthesis and uptake in animal cells. Peptides which bind to and are
CC cleaved by SKI-1 may be used for monitoring SKI-1 activity, for screening
CC inhibitors of SKI-1 activity, or for screening enhancers of SKI-1
CC activity. Proteic fragments of SKI-1 which bind to the SKI-1
CC catalytic site may be used as inhibitors of SKI-1 activity. They may
CC be used to treat diseases involving overexpression of SKI-1 or SKI-1
CC substrate. Such diseases include hypercholesterolemia, high levels of
CC fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,

QY 701 GluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu 720
DB 2108 GAGTACTCCCTGAGGAGATGCTAGCTGAGGAGGATGTGGCAATGGCCCTTCCCTC 2167
QY 721 ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
DB 2168 GTCATCTTCAGTGACTGCTACAACTCTCTGTTATGAGAAAGTGAAGTTTATGATGAA 2227
QY 741 AsnThrArgGlnTrpTrpMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu 760
DB 2228 AACACACGAGCTGGTGGATGCGACACCGGAGGAGCGCAATCCAGCTCTGAATGAG 2287
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
DB 2288 CTGCTGCTGTGTGGAACTATGGGCTTCAGTGACGGCTTATGAAAGGAGTTGTCCG 2347
QY 781 AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
DB 2348 GCAAAACCATGACATGCTATGCTGGGTGCGAGCATGCCAAGTTTCCAGAAGATGGC 2407
QY 801 ValValIleThrClnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
DB 2408 GTGCTGATCACACAGACTTTCAGGACCAAGGATGGAGTCTTAAACAAGACAGCA 2467
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
DB 2468 GTTGTGGAATGTTCCCTATTTGGGGCTTATCAGATTCCATCTGAAGTGGAGCGCG 2527
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
DB 2528 ATCGTCTGTATGAGACTCCACTGCTTGGATGACAGTCACAGACAGAGGACTGCTT 2587
QY 861 TrpLeuLeuAspAlaLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
DB 2588 TGGCTTCTGATGCGTCTCTAGTACATCTATGGCGTGACCTCCACCCCTCAGC 2647
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
DB 2648 CATTCAGGAAACCGGAGCGCCACCTAGCGGAGCGGCTGGCCCTCTCTGAAGGATG 2707
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
DB 2708 GAAGGAAACACCTCCATCGGTACTCCAAAGTTCTTGAAGCCCACTTGGAGACCGGAA 2767
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
DB 2768 CTTGGCCCTGCGGAGCTGTCACATTTGTCATGGGCCAAGCCACAGCCTTTGAATGAG 2827
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
DB 2828 ACGGCACCCAGTAATCTTTGGAACATCAGAGCTGCTCTCCATTACCTGGCAAGTA 2887
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
DB 2888 GTGTTACCAACTTTCATCCATCGCCCTCAAGTGAGAGACTTGTCCCTCGAGAGAGT 2947
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
DB 2948 GTGCTCTGGACATCTCTGGAGGATCATGCTTGGCCCTTACAAACAGGAGTGGGACAG 3007
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
DB 3008 ACCATCCCGCTCTTCGCTTCTCGGAGCATGCTGGCCCTGCTTGTGGGTACAG 3067
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
DB 3068 ATCAGCAAGCCCAAGACCGCGGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3127
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
DB 3128 GCACAGCAGGCGCCACCTTCGACAGGACCCCATCATG 3163

RESULT 6

ABL90664/c
ID ABL90664 standard; cDNA; 2729 BP.

XX AC ABL90664;

XX DT 24-MAY-2002 (first entry)

XX DE Human polynucleotide SEQ ID NO 1226.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW antiallergic; hepatotrophic; antidiabetic; antiinflammatory; antiulcer;
XX KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein; gene; ss.

OS Homo sapiens.

XX PN WO200190304-A2.

XX PD 29-NOV-2001.

XX PF 18-MAY-2001; 2001WO-US16450.

XX PR 19-MAY-2000; 2000US-205515P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Birse CE, Rosen CA;

XX WPI; 2002-122018/16.

DR P-PSDB; ABB90255.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
prevention of neural, immune system, muscular, reproductive,
gastrointestinal, pulmonary, cardiovascular, renal and proliferative
disorders -

XX Claim 4; SEQ ID NO 1236; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL90449-ABL90853) and proteins
(ABB9040-ABB90444) useful for preventing, treating or ameliorating
medical conditions e.g. by protein or gene therapy. The genes are
isolated from a range of human tissues disclosed in the specification.

XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
and ovarian cancer and other cancers of the adrenal gland, bone, bone
marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
disease, multiple sclerosis, rheumatoid arthritis and ulcerative
colitis; (c) cardiovascular disorders such as myocardial ischaemias;

XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
and parasitic infections.

XX Note: The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at wipo.int/pub/published_pct_sequences.

XX Sequence 2729 BP; 642 A; 722 C; 674 G; 686 T; 5 other;

Alignment Scores:

Pred. No.: 6,72e-241 Length: 2729
Score: 3560.00 Matches: 665
Percent Similarity: 99.55% Conservatives: 0
Best Local Similarity: 99.55% Mismatches: 3
Query Match: 63.38% Indels: 0
DB: 24 Gaps: 0

US-09-830-837-6 (1-1052) x ABL90664 (1-2729)

QY 385 GlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGlySerGlyValLys 404
|||||

animal model; lipid metabolism; transcription factor; cholesterol; obesity; insulin resistance; therapy; diagnosis; pesticide; ss.

Drosophila melanogaster.

Key Location/Qualifiers
CDS 62..3040
/*tag= a

WO200076308-A1.

21-DEC-2000.

08-JUN-2000; 2000WO-US15880.

14-JUN-1999; 99US-0332522.

15-MAR-2000; 2000US-0189700.

(EXEL-) EXELIXIS INC.

Costa MA, Doberstein SK, Elson S, Ferguson KC, Homburger SA;
Ebels AJ, Keegan KP, Stout TJ;

WPI; 2001-091292/10.

P-PSDB; AAB20015.

Novel invertebrate organism genetically modified to express or
mis-express steroid regulatory element binding protein pathway protein
used as model system for studying lipid metabolism and determining
lipid content

Claim 9; Page 79-80; 90pp; English.

The present sequence is that of *Drosophila melanogaster* DNA
encoding dSIP (see AAB20015), a homologue of mammalian site 1
proteases (SIP) that are involved in sterol regulatory element
binding protein (SREBP) processing, cleaving SREBP at the luminal
loop. SREBPs are transcription factors that activate genes
involved in cholesterol and fatty acid synthesis, and are major
mediators of insulin action in the liver. Alterations in SREBP
function and expression are implicated in obesity and insulin
resistance. The dSIP DNA was identified from an expressed sequence
tag database search on the basis of homology to hamster S2P. The
invention provides fly and nematode (*Caenorhabditis elegans*)
polynucleotides (see AAB9287-90) and encoded proteins (see AAB20012-15)
involved in the SREBP pathway. *C. elegans* and *D. melanogaster*
animals genetically modified to express or mis-express these
proteins are claimed. These genetically modified animal models have
identifiable phenotypes that make them useful in assays for studying
lipid metabolism, other genes implicated in lipid metabolism and
compounds capable of modulating lipid biosynthetic pathways. Model
organisms or cultured cells can be used to identify new drug
targets, therapeutic agents, diagnostics and prognostics of
disorders associated with lipid metabolism, and also to identify
pesticide targets directed to components of the SREBP pathway.

Sequence 3160 BP; 985 A; 671 C; 667 G; 835 T; 2 other;

Alignment Scores:

Pred. No.:	7.53e-157	Length:	3160
Score:	2366.50	Matches:	466
Percent Similarity:	61.70%	Conservative:	151
Best Local Similarity:	46.60%	Mismatches:	232
Query Match:	42.13%	Indels:	151
DB:	22	Gaps:	16

US-09-830-837-6 (1-1052) x AAB9290 (1-3160)

QY	47	PheSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAla	66
Db	116	TTTAAACAGCGCTGTTCCTCAATGAGTTTCATCTCCACTTCCATCAAAATCTTGCC	175
QY	67	LysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArg	86

Db	176	CGGTCGCGAATCCTACATCGCAGCAAAATCTTGGTTCAACAGTAACGAACACTGGAGA	235
QY	87	IleIleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIle	106
Db	236	ATTGTTCCCGCTCTAAATTTGGCTTGGCAATATCAAGTAGTATTTGATATCTTACGAGTT	295
QY	107	LysGluLysGlnLysAlaGly-----LeuLeuThrLeuGluAspHisProAsn	122
Db	296	TGCGACGGTTATGAATCATCATCAGAGTTTATTATAGAAAGGCTTCAGACTCACCATCA	355
QY	123	IleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyr---AlaGluSer	141
Db	356	GTAAGGCGAGTAGTTCCCGCAGCGAAGGATCCTAAACTATGACGCTATAGC	415
QY	142	AspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgPro	161
Db	416	AACCTAACG-----	424
QY	162	LeuArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSer	181
Db	425	-----TATATTAC-----CGCCATCCC	442
QY	182	SerArgArgLeuLeuArgAlaIlePro-----ArgGlnValAlaGlnThr	196
Db	443	CAAGGAGTGTCTAAGGAACAGAAACCCCAACAGATCGCCACCGACAATTTGTCTCGTA	502
QY	197	LeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaVal	216
Db	503	CTCCAGCCCAACATCTTTTGGAACTGGGTATCACAGCAAGGAGTAAAGTGGCCATT	562
QY	217	PheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsn	236
Db	563	TTGCACACTGGCCTAACCAAAACCATCCACACTTTCGAATGTAAAGAACCAACAAAC	622
QY	237	TrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyVal	256
Db	623	TGCACGAATGAAAGTCACTTGACGACAGAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG	682
QY	257	IleAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArg	276
Db	683	ATCGCTCTCTCCAGGAATGCCCTAGCTTCGCTCCGACGCCCTTTTACATATTATAA	742
QY	277	ValPheThrAsnAsnGlnValSerTyrThrSerTyrPheLeuAspAlaPheAsnTyrAla	296
Db	743	GTITTTACGACACTCCCAAGTTCTTACACTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG	802
QY	297	IleLeuLysLysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHis	316
Db	803	ATATATAGGAAAAATAAATCTCAACCTTAGCATTTGGGGTCCCGACTTTATGGACTCG	862
QY	317	ProPheValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIle	336
Db	863	CCGTTCTGTTGAAAGGTGTGGAACTGTCGGCTGTAATAATGCTAATGATATATGCGGCA	922
QY	337	GlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIle	356
Db	923	GGAAATGATGTTCCCTTTGACGCGACGCTAAACAACTCTGGCGCATCAGACGATGATGT	982
QY	357	GlyValIleGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThr	376
Db	983	GGCGTTGGTGGCATTCAGTTTGTATGATAAAATCGCCCAAGTTTAGTTCGAGAGGAATGAC	1042
QY	377	ThrTrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAla	396
Db	1043	ACGTGGGAACCTTCCCTTAGCTACGACGCTATGGGACTCGATATTCACGTACGGAAGT	1102
QY	397	GlyValArgGlySerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla	416
Db	1103	CAAGTGGAAAGCGAGTAGTATGCGCAAGGGGTGCAGACGACTCTCTCGAACATCCGTTGC	1162
QY	417	SerProValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGlu	435

Best Local Similarity:	40.10%	Mismatches:	271
Query Match:	39.51%	Indels:	287
DB:	23	Gaps:	21
US-09-830-837-6 (1-1052) x ABL12460 (1-5332)			
QY	47	PheSerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAla	66
DB	4278	TTTAAACAGCCGTTGTCCTCAATGAGTTCATCGTTCACTCCATCAAAATACTTTGCC	4219
QY	67	LysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluVal	82
DB	4218	CCGGTCCGAGATCCCTACATCGCAGCAAACTTCTTGGTCAACGTAAGTTGTATTCA	4159
QY	82	-----	82
DB	4158	AATGGTTATATTATACCTTAAATGTCATTCTCCTGTTTATTTCAACGGATTGAGCCT	4099
QY	82	-----	82
DB	4098	GCTTCTTTGAAAGACACAAAGCATCTTTTCTGCAGAGAAAGTACCTTTTAAAGTTT	4039
QY	83	-----	83
DB	4038	AAACACTTTTCAACAGCTCTTTATATCCCTTTCCATTTTATAGTAACGAATCGGAGAT	3979
QY	87	eIleProArgAsnAspProSerSerAspTyrProSerAspPheGluValIleGlnIleLys	107
DB	3978	TGTTCCCGCTTAAATTTGGCTGGCAATATCAAGTATTTTGTATCTTACGAGTTTG	3919
QY	107	sGluLysGlnLysAlaGly	123
DB	3918	CGACGGTTATGATCATCATCATGAGGTTTATTATAGAAAGCTTCAGACTCACCCACT	3859
QY	123	eLysArgValThrProGlnArgLysValPheArgSerLeuLysTyr	142
DB	3858	AAGGCAGTAGTTCCTCCAGGAGCTACGAGGATCCTTAAACTATGACGCTATAGCAA	3799
QY	142	pProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerArgProLe	162
DB	3798	CCTAAGC	3792
QY	162	uArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSe	182
DB	3791	-----TATATTAC-----CGCATCCCA	3772
QY	182	rArgArgLeuLeuArgAlaIlePro	197
DB	3771	AGGAGTGTAAAGAACAGAAACCCAAACACGATCGCCACCAATTTGCTCCGTACT	3712
QY	197	uGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPh	217
DB	3711	CCACGCCACATCCTTTGGAAGCTGGTATACAGCGAAGGAGTTAAAGTGGCCATTTT	3652
QY	217	eAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTr	237
DB	3651	CGACACTGGCTTAAACAAAACCATCCACACTTTCGAATGTAAAGGACGAAACAACTG	3592
QY	237	pThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValI	257
DB	3591	GACGAATGAATTCACCTTCGACAGACAGATCATGGCACCTTCGTCGCGGGGTAAT	3532
QY	257	eAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgVa	277
DB	3531	CGCTTCTCCAGGATGCTAGGCTTCCTCCGACGCCGATCTTACATATTAAAGT	3472
QY	277	lPheThrAsn	280
DB	3471	TTTTACGAACCTCCCAAGTGAGCAACGTAAGTGAATTAACCTAAGCAATCTAATTAG	3412
QY	281	-----Asn-GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleL	298
DB	3411	CCCTACATCGGCGAGGTTTCTTACCTTCTGCTTCTCGATGCTTCACTACGCGATAT	3352
QY	298	eulLysLysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProp	318
DB	3351	ATAGGAAATAACATCTCAACCTTAGCATGGGGTCCGACTTTTATGGACTCCCGCT	3392
QY	318	heValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyA	338
DB	3291	TCGTTCAAAAGGTTGGAACTGTCGCCTAATAATGTCATAATGATATCGGCAGGAA	3332
QY	338	snAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyV	358
DB	3231	ATGATGCTCCCTTGTAGCGCACGCTAAACAATCCTGCGCATCAGACGATGTATGTGGC	3172
QY	358	aIlglyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr	378
DB	3171	TTGGTGGCATTCAGTTTGTATGATAAATCGCCAAAGTTTAGTTCGAGAGNATGACACGT	3112
QY	378	rpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyV	398
DB	3111	GGGAACCTTCCCTTAGGCTACGGACGTATGGACTCGATATTGTGACATCCGTCCTCTC	3052
QY	398	aIArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerP	418
DB	3051	TGAAGGCGAGTGATGTCGCAAGGGTGCAGACGACTCTCTGGACATCCGTCTCTCTC	2992
QY	418	roValValAlaGlyAlaValThrLeuLeuValSer	437
DB	2991	CAGTTGTCGAGGGGCTGCTGCACCTGCTTATAGCGGTCATTTAGAAAATCGACTACA	2932
QY	437	aIAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValA	457
DB	2931	TAAACCCAGCATCTCTTAAAGCAGGTACTCATTAAGGTCGCGAGAAACTGCCGCA	2872
QY	457	snMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnS	477
DB	2871	ACATGTTTACGAGGAGCTGGAAGTCTGCTGAAGATATGCAAGTATTCGACTATTCTGT	2812
QY	477	eTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrM	497
DB	2811	CATACAAACCAAGATAACCTTATTCGCGCATACCTTCACCTCACCCAA--AACTATA	2755
QY	497	eTTrpProTyrCysSerGlnProIleTyrTyrGlyMetProThrValValAsnValT	517
DB	2754	TGTGGCTTATAGTCCCAACCTCTGCTACTTGAAGCTCCGTCGCTATTGCAACGTTA	2695
QY	517	hrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProT	537
DB	2694	CCATCTCAATGGTATCTCTGTCAAGTCATATAGTTGGCATCCCTCAATGGATTCGCG	2635
QY	537	yrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpP	557
DB	2634	ATTTCGAAACCAAGTCTAGTTCTTCAAGTATCTGCACAAAGTTCCGCTATCGTTTGGC	2575
QY	557	roTTrpSerGlyTyrLeuAlaIleSerIle	566
DB	2574	CGTGGACCGTTGGATGTCAGTTTATTTGGTGAGTATGATAAATAGCAACGCTATATAA	2515
QY	566	-----	566
DB	2514	AGGATATATGGGTACATATTGATATGAAGTGTAAAGCCATTCTATATACCAACAATTTT	2455
QY	567	-----ValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisVal	583
DB	2454	TGTTTACCAGCTGTAAGGAGGAGAAACTTTGAAGGTTTGTATAGGAAGATATC	2395
QY	584	MetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThrSer	603
DB	2394	ACCTTAGTTTGGAAAGCTTTAAACAGACACCAACCAAGAACT-----CATGTTACA	2344
QY	604	ThrValLysLeuProIleLysValLysIleIleProThrProProArgSerLysArgVal	623
DB	2343	GAAGTCGACTTCTTCTTACATAAAGGTTACTCCAAACCCGCAAGAAAGAGATT	2284

```
QY 624 LeuTyrAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArgAspAsnLeu 643
|||||
2283 TTATGGGATCAGTACCACAGCCTTAGGTATCCACCGCTATATCCACGAGATGATCTC 2224
QY 644 ArgMetLysAsnAspProLeuAspTyrAsnGlyAspHisIleHisThrAsnPheArgasp 663
|||||
2223 AAAGTTAACTAGATCTCTGGACTGGAGGCGAGACATATACACACAACTTTAGGAC 2164
QY 664 MetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThr 683
|||||
2163 ATGTATACACATTTACGAAATGTTGGCTACTACATGATGTTTGGCAGAACCTTCACC 2104
QY 684 CysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGlyPhe 703
|||||
2103 TGCCTCATGCTCGGATTAAGCGCTTATGATGTTGACCTGAGAGAGGTTTGGC 2044
QY 704 ProGluGluIleAlaLysLeuArgAspVal---AspAsnGlyLeuSerLeuValIle 722
|||||
2043 GACGAGGAATAAACGCTTTACGGAACAGCTGTATAAAGAGCGTTGAATGTCGCGTA 1984
QY 723 PheSerAspTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThr 742
|||||
1983 TTCGAGACTGGTATACACACCTGTGATGAAAAAATTAATCTTTGACGAGAACACC 1924
QY 743 ArgGlnTyrPheMetProAspThrGlyAlaAsnIleProAlaLeuAsnGluLeuLeu 762
|||||
1923 CGACAAATGGTGACACCCGACACTGGTGGCGCAATATTCAGCCTTGAATGATTATG 1864
QY 763 SerValTyrAsnMetGlyPheSerAspGlyLeuTyrGlyGluPheThrLeuAlaAsn 782
|||||
1863 AAGCCATTTGAATTTGCGCGATTTTTCGGTGGAGGACATTTCAAACTGGCGGAC 1804
QY 783 HisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAsp-----Gly 800
|||||
1803 CATTCAATGACTACTGTAGTGGAGCCACAAATGTTAAGTTTCCAATGAATCCAGGAGAT 1744
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluVal----- 814
|||||
1743 ATTATAGTGGGCAAAACAACTGAATGACCAAGGACTTTGGTAAATTTAATAATCCTTT 1684
QY 815 -----LeuLysGlnGluThr 819
|||||
1683 TTAGTTTCCAGAAATACAACTTTGCCTTTTCCAGATTAATTAATCTAANAACCC-AGC 1625
QY 820 AlaValValGlu---AsnValProIleLeuGlyLeuTyrGlnIleProAla----- 835
|||||
1624 AAGGTAGCAAACTAGATCTACCTATTTTGGTGTGTTCCAAACCAAGGCGAACAGTATT 1585
QY 835 ----- 835
1564 CAACAGCAGGAGAAATCGTGCAATGCGGAAAGCAATTTGGCAGAGGCTATACCACA 1505
QY 835 ----- 835
1504 GATTACTCCACATTTAAGAACCGGTTTTTGTCTACTCGCAACGAAACGAAGTATCAGT 1445
QY 836 -----GluGlyGlyArgIleValLeuTyrGlyAsp 846
|||||
1444 TTTGCGAAAAGCAATATCATGAACCTAAGAATGAAGACGCTATTGCGGTATATGGGAC 1385
QY 847 SerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeu 866
|||||
1384 TCCAACTGCCTCCGACTCCACGCACTGAGAGGCTTGTCTACTGCTGTCTAATAAGTTT 1325
QY 867 LeuGlnTyrThrSerTyrGlyValThrProProSerLeuSerHisSerGlyAsnArgGln 886
|||||
1324 TTAGATTTCATAAATAC-----TCGAC----- 1301
QY 887 ArgProProSerGlyAlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHis 906
|||||
1300 -----AAATCAGTTTATGCGAG 1293
QY 907 ArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLysProArgProLeuProAla 926
```

RESULT 10

```
AAAX39701
ID AAX39701 standard; DNA; 786 BP.
XX
AC AAX39701;
XX
DT 02-JUL-1999 (first entry)
XX
DE Gastric cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN W09904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfeundschtuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 516; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
```

CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HIA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.

XX Sequence 786 BP; 202 A; 190 C; 214 G; 172 T; 8 other;

Alignment Scores:
 Pred. No.: 1,266-77 Length: 786
 Score: 1229.50 Matches: 239
 Percent Similarity: 91.92% Conservative: 0
 Best Local Similarity: 91.92% Mismatches: 18
 Query Match: 21.89% Indels: 4
 DB: 20 Gaps: 1

US-09-830-837-6 (1-1052) x AAX39701 (1-786)

QY 692 ThrLeuLeuMetValAspSerGluGluGluTyrPheProGluGluLeuAlaLysLeuArg 711
 DB 3 ACTTCTGATGTTGACAGTGGAGGAGGAGTACTTCCCTGAAGAGATGCCAAGCTCCGG 62
 QY 712 ArgAspValAspAsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThrSerVal 731
 DB 63 AGGAGCTGGACAAACCGCTCTCGCTCGTATCTTCACTGCTGTTACAACTCTCTGTT 122
 QY 732 MetArgLysValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGly 751
 DB 123 ATGAGAAAGTGAAGTTTATGATGAACAAACAGGAGGAGTGGATGCCGGATACCGGA 182
 QY 752 GlyAlaAsnIleProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAsp 771
 DB 183 GGAGCTAAATCCACGCTCTGAATGAGTGTCTGTGTGTGGAACATGGGTTTCAGCGAT 242
 QY 772 GlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTrpAlaSerGlyCys 791
 DB 243 GGCTGTATGAAGGGAGTTTCACTCCCTGGCCCAACATGACATGATTTATCGCTCAGGTCG 302
 QY 792 SerIleAlaLysPheProGluAspGlyValValIleThrGlnThrPheLysAspGlnGly 811
 DB 303 AGCATCGGAAGTTTCCAGAGATGGCTGCTGATTAACAGACTTTCAGGACCAAGGA 362
 QY 812 LeuGluValLeuLysGlnGluThrAlaValAlaValGluAsnValProIleLeuGlyLeuTyr 831
 DB 363 TTGGAGGTTTAAAGCAGGAACAGCAGTGTGTTGAAACGCTCCCATTTTGGGACTTAT 422
 QY 832 GlnIleProAlaGluGlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAsp 851
 DB 423 CAGATTCCTAGCTCAGGAGGAGGAGGAGTGTACTGTATGGGAGCTCCCAATTCCTGGAT 482
 QY 852 AspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuGlnTyrThrSer 871
 DB 483 GACAGTCACGACAGAGAGGACTCTTTTGGCTTCTGGTGGCTCTCCAGTACACATCG 542
 QY 872 TyrGlyValThrProProSerLeuSerHisSerGlyAsnArgGlnArgProProSerGly 891
 DB 543 TATGGGTGACACCGCTAGCTCAGTCACTCTGGGAACCGCCAGCGCCCTCCANTTGG 602
 QY 891 yAlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHisArgTyrSerLysVal 911
 DB 603 AGCAAGCTCAGTCACTCCAGAGAGGATGGAAGAAACCATCTCTCATCGGTACTTCCAAGGT 661
 QY 911 lLeuGluAlaHisGlyAspProLysProLysProArgProLeuProAlaCysProArgLeuSe 931
 DB 662 TCTGGAGGCCATTTGGGAACCAACCAACCTCGGGCTCNACACCTGTCACGCTGTN 721
 QY 931 rTTPAlaLysPro-GlnProLeuAsnGluThr---AlaProSerAsnLeuTrpLys 948

Db 722 CTGGGCCAAGCCCAANAGCCTTAAACCAANACGCGCCCAATTAACCTTTGGAAA 777
 RESULT 11
 AAZ15839
 ID AAZ15839 standard; cDNA; 757 BP.
 XX AAZ15839;
 AC AAZ15839;
 DT 12-OCT-1999 (first entry)
 XX Human gene expression product cDNA sequence SEQ ID NO:3308.

XX Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX Homo sapiens.
 XX WO9938972-A2.
 XX 05-AUG-1999.
 XX 28-JAN-1999; 99WO-US01619.
 XX 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080315.
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.

PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX WPI; 1999-494092/41.

Novel human genes and their expression products which are
 differentially expressed in different cell types

Claim 1; Page 1586; 2479pp; English.

The present invention describes a library of human polynucleotides
 comprising the sequences given in AAZ12532 to AAZ17779. Also described is
 a method of detecting differentially expressed genes correlated with the
 cancerous state of a mammalian cell, comprising detecting at least one
 differentially expressed gene product in a test sample from a cell
 suspected of being cancerous, where the gene product is encoded by one
 of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
 polynucleotides can be used as a source of primers and probes, which can
 be used for a variety of purpose, e.g. detection of expression levels,
 mapping, tissue typing or profiling, forensics, genetic analysis and
 detection of polymorphisms. Polypeptides encoded by the polynucleotides
 can be used for raising antibodies for experimental, diagnostic and
 therapeutic purposes. The polynucleotides may also be used to construct
 arrays for diagnostics (which may be used to determine function of an
 encoded protein); and to detect differences in expression levels between
 two cells (e.g. to identify abnormal or diseased tissue in a human, to
 identify a genetic predisposition or susceptibility to a disease such as
 cancer). The polynucleotides of the invention are especially used in the
 diagnosis, prognosis and management of colorectal cancer, breast cancer,
 and lung cancer. The polynucleotides can also be used to screen for
 peptide analogues and antagonists.

Sequence 757 BP; 157 A; 198 C; 187 G; 161 T; 54 other;

Alignment Scores: 2.09e-72 Length: 757
 Pred. No.:

CC proteins, antibodies and (ant)agonists are useful in the diagnosis,
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
CC and other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 9689 BP: 2184 A: 2448 C: 2449 G: 2608 T: 0 other:

Alignment Scores:	
Pred. No.:	3,21e-14
Score:	348.00
Percent Similarity:	49.5%
Best Local Similarity:	45.50%
Query Match:	6.20%
DB:	22
Length:	9689
Matches:	101
Conservative:	9
Mismatches:	52
Indels:	60
Gaps:	9

US-09-830-837-6 (1-1052) x AAI63568 (1-9689)

Qy	875	ThrProPro	---SerLeuSerHisSerGlyAsnArgGlnArgProSerSerGlyAlaGly	893
Db	8380	AGCCCCCTTCCTCCCTCCCACTCA	-----GCTGGA	8412
Qy	894	SerValThrProGlnArgMetGluGlyAsnHisLeuHisArgTyrSerLysValLeuGlu	913	
Db	8413	ACCACAGCCCCCTTCCTCCCTCCGTAGAGAGCTGCCCATCTTCCTCGTTGTTTAGAG	8472	
Qy	914	AlaHis	-----LeuGlyAspProLys	920
Db	8473	AAACACAGACGCGCTCTGGCGCTCCCTCGGTAGTGTGTGACAGAACATAGCTGGCGCT	8532	
Qy	921	ProArgProLeuProAlaCysPro	---ArgLeuSerTrpAlaLysProGlnProLeuAsn	939
Db	8533	GATCATCCGCTCTCGCGGGATCCCGAGGAGGAGGCGACAGAGATCCACCGAGCCTCCAC	8592	
Qy	940	GluThrAla	-----ProSerAsnLeuTrpLysHisGlnLysLeuLeu	953
Db	8593	CCCACGAAAGTGGCTCTTTGGCTCTCATCCACTCTTTATTATATAGCTGGGTACTT	8652	
Qy	954	SerIleAspLeuAspLysValValLeuPro	-----	963
Db	8653	TAGCTAAGGCTTAAAAATGATGATCCCTCCAGCCAGCTAGTGAAATAGCTCGCAT	8712	
Qy	964	-----AsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSerGly	981	
Db	8713	TTAGTTTCTGATCAATCCAGCCGCTCATAGAAAAAGCGC	8751	
Qy	982	AlaTrpAspIleProGly	-----GlyIleMet	990
Db	8752	TCATGTTAGCTGTGGGGTTCCTGCTTCAACCGGCTGCTCTTCACAGGATCATG	8811	
Qy	991	ProGlyArgTyrAsnGlnValGlyGlnThrIleProValPheAlaPheLeuGlyAla	1010	
Db	8812	CCTGCCGCTACACAGAGGTGGCCAGACCATCTCTGTCTTTCCTCCGGAGCC	8871	
Qy	1011	MetValValLeuAlaPhePheValValGlnIleAsnLysAlaLysSerArgProLysArg	1030	
Db	8872	ATGGTGTCTTCGGCTCTTTGTGGTACAAATCAACAGGCCAAGCAGCGCCGAAGCGG	8931	
Qy	1031	ArgLysProArgValLysArgProGlnLeuMetGlnIleValHisProPolysThrPro	1050	
Db	8932	AGGAAGCCAGGTTAAGCCCGCAGCTCATGTGACGAGGTTTCAACCGCCCAAGACCCCT	8991	
Qy	1051	SerVal	1052	

Db 8992 TCGGTG 8997

RESULT 15
AAN80745

ID AAN80745 standard; DNA; 1220 BP.

XX AAN80745;

XX 14-SEP-1990 (first entry)

XX Coding region of the aprA gene encoding subtilisin in *Bacillus subtilis*.

XX *Bacillus subtilis* strain Q8127 (trpC2 leuA8 sacuH200); subtilisin;
XX thermal stability; pH stability; specific activity;
XX substrate specificity; detergency; ss.

XX *Bacillus subtilis* strain Q8127 (trpC2 leuA8 sacuH200).

OS

XX Key Location/Qualifiers

FF mat_peptide 1..1143

FT /*tag= a

FN W08808033-A.

PN

PD 20-OCT-1988.

PP 28-MAR-1988; 88WO-US01038.

PR 10-APR-1988; 88US-0036872.

PA (AMGE-) AMGEN INC.

XX Zurewski MM, Stabinsky Y, Levitt M;

PI WPI: 1988-307568/43.

DR P-PSDB; AAP80744.

XX New subtilisin analogues -
XX have aminoacids(s) present in calcium binding site replaced by
XX negatively charged aminoacid(s)

XX Example 2; Table 1, Pages 22-24; 60pp; English.

XX Genomic DNA was isolated from cells of *B. subtilis* strain Q8127 (trpC2
XX leuA8 sacuH200). The coding region of the aprA gene was sequenced and
XX the results of the sequence are given in AAN80745. The specific
XX identity of the initial 5 codons of the leader region is attributable to
XX the report of Stahl, et al., J. Bacteriol., 158, 411-418, (1984) and
XX Wong, et al P.N.A.S., 81, 1184-1188 (1984). There exist codon sequence
XX differences from Stahl, et al., at amino acid positions 84 and 85.
XX Specifically, Stahl, et al., reports a codon GTA (coding for valine) at
XX amino acid position 84 while the codon GTA (also coding for valine)
XX appears in AAN80745. Stahl, et al., also reports a codon AGC (coding
XX for serine) at amino acid position 85 as opposed to the codon GCG (coding
XX for alanine) in AAN80745. The patent concerns a novel subtilisin
XX analogue which has an amino acid sequence of a naturally occurring
XX *Bacillus subtilis* in which has been modified by having: one or more of the
XX amino acids present in a calcium binding site of the naturally occurring
XX *Bacillus subtilis* replaced by a negatively charged amino acid, and
XX one or more of any Asn-Gly sequence of the naturally occurring *Bacillus*
XX subtilisin deleted or replaced by a different amino acid. Pref. it is
XX an analogue of subtilisin Carlsberg, subtilisin Df, subtilisin BPN', an
XX aprA subtilisin of *B. subtilis* or subtilisin from *B. mesentericus*. The
XX subtilisin analogues exhibit improved thermal and pH stability,
XX increased specific activity and broad substrate specificity thereby
XX increasing the detergency of detergent formulations contg. such
XX analogues.

XX Sequence 1220 BP; 355 A; 281 C; 283 G; 301 T; 0 other;

XX

Alignment Scores: 7.12e-15 Length: 1220
Pred. No.: 341.00 Matches: 125
Score:

DR P-PSDB; AAP60571.

XX Bacillus strains with reduced extra-cellular protease levels -
 PT useful as hosts for secretion of heterologous polypeptide(s) and
 PT proteins.

XX Example; Fig 2; 30pp; English.

XX In the example, the inventors inactivate the apr gene by insertion
 CC of a functional cat gene coding for chloramphenicol
 CC acetyltransferase from *S. aureus*. The inserted cat gene confers
 CC chloramphenicol resistance, thus facilitating the selection of
 CC transformants. The reduced levels of subtilisin greatly reduces the
 CC likelihood that a secreted foreign protein will be proteolytically
 CC degraded prior to recovery.

XX Sequence 1524 BP; 461 A; 329 C; 331 G; 403 T; 0 other;

Alignment Scores:

Pred. No.:	9,46e-15	Length:	1524
Score:	341.00	Matches:	125
Percent Similarity:	41.96%	Conservative:	68
Best Local Similarity:	27.17%	Mismatches:	184
Query Match:	6.07%	Indels:	103
DB:	7	Gaps:	19

US-09-830-837-6 (1-1052) x AAN60475 (1-1524)

QY	48	SerSerThrValValGluThrGluThrIleValAlaPheAsnGlyTyrPheThrAlaLys	67
DB	379	AGCATAC-----GAAAGAAATACATGTCGGATT-----AAA	414
QY	68	AlaArgAsnSerPheIleSerAlaLeuLysSerSerGluValAspAsnIrpArgIle	87
DB	415	CAGACAAAGAGTCCCATGATTCGCCCAAGAAAGGATGTTAT-----	459
QY	88	IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys	107
DB	459	-----	459
QY	108	GluLysGlnLysAlaGlyLeuLeuThrLeuGlu-----AspHisProAsnIleLysArgVal	126
DB	460	-----TCTGAAAGAGCGGAAGGTTCAAAGCAATTTAAGTATGTTAACCGCGCCGACGA	516
QY	127	ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro	146
DB	517	ACATTGGATGAAGAAGCTTAAGAATTTGAA-----AAGATCCGAGCGTTGCA	567
QY	147	CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer	166
DB	567	-----	567
QY	167	LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeu	186
DB	568	-----TATGTGGAAGAGATCATATTGCACATGAATATGCG	603
QY	187	ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly	206
DB	604	CAATCTGTCTTATGGCATTTCTCAA---ATAAAGCGCCGCTCTTCACCTCTCAAGC	560
QY	207	TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro	226
DB	661	TACACAGGCTCTAACGTAAGTAGCTGTTATCGACAGCGGAATTGACTCTTCTCATCCT	720
QY	227	HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp	244
DB	721	GACTTA---AAGCTCAGAGCGGAGCAAGCTTCGTACCTTCCTCAACCAACCAATACCA	777
QY	245	AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----	262
DB	778	GACGGAGTCTCAGCGGTACGATGATGCGGTAGATGCGGCTTCTTAATCAATCAATC	837
QY	263	---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn	281

DB	838	GGTGTCTTGGCGTAGCGCAAGCGCATCATATATATGTCAGTAAAGTCTGTGATTCACA	897
QY	282	GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysIle	301
DB	898	GGAAGCGGCAATATAGCTGGATTATTAACGCAATTGAGTGGCCATTTCCACAAATATG	957
QY	302	AspValLeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheVal	319
DB	958	GATGTATACATAGACCTTGGCGACCTACTGTTCTACAGCGCTGAACACAGTCGTT	1017
QY	320	AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp	339
DB	1018	GACAAAGCC-----GTTTCCAGCGGTATGCTGCTGCTCCGACGCGGAACGAA	1068
QY	340	GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly	357
DB	1069	GGTTCATCCGGAAGCACAGCAGTCGCTACCTGCAAAATATCTCTACTATATGCA	1128
QY	358	ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr	377
DB	1129	GTAGTGGCGTAAACAGACGACCAAGAGCTTCAATCTCCAGCGCAGGTCTCT-----	1182
QY	378	TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly	397
DB	1183	---GAGCTT-----GATGTGATGCTCTCGCGGTGTC	1212
QY	398	ValArgGlySerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer	417
DB	1213	ATCCCAAGCACACTTCTCGAGCGCTTACGCGCTTATAACGGAACGTCCTACGCGACT	1372
QY	418	ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal	437
DB	1273	CCTCAGTTGCGGAGCAGCAGCGTTAATCTTCT-----AAGCACCGACTTGG	1323
QY	438	AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn	457
DB	1324	ACAAAGCGCAAGTCGCTGATGCTTTAGAAAGCACTGCAACATATCTTGA-----AAC	1377
QY	458	MetPheGluGlyHisGlyLysLeuAspLeuLeuArgAla-TyrGlnIleLeuAsnSe	477
DB	1378	TCTTCTACTATGGAAGGGTTAATCAACGTACAAAGCAGCTGCACAAAT-----AATAG	1431
QY	477	rTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyr	496
DB	1432	TAAAGAGACGAGTT-----CCTCCATACCTGCTCTTTTATTGTCAGCAT	1480

RESULT 18

AAN81505
 ID AAN81505 standard; DNA; 2044 BP.

XX AAN81505;

XX 10-SEP-1990 (first entry)

XX DNA encoding the alkaline protease gene promoter domain, and domain
 DE relating to protein secretion of *Bacillus licheniformis*.

XX Alkaline protease gene promoter domain; *Bacillus licheniformis*;
 KW *Bacillus subtilis* PW10; alkaline protease purification.

XX *Bacillus licheniformis*.

Key	Location/Qualifiers
Promoter	149..150
promoter	/*tag= a
promoter	151..154
promoter	/*tag= b
promoter	172..177
promoter	/*tag= c
promoter	209..213
repeat_unit	/*tag= d
	1389..1400

Db 1391 AAAGCTAGTGTGTTTACGACTAGCTTTTCTTCATTCAGTTGAA-CACGTGTTCAATATTT 1449
 QY 498 -----TTPProTyrCysSerGlnProIleTyrGlyGly 509
 Db 1450 TGAATCCGTTCCATATGTCGGATGCGCGGTATTTAAATAATCTTGACGAGAACGCGGG 1509
 QY 510 MetProThrVal 513
 Db 1510 TTCGCTCGCTC 1521

RESULT 19

AAAN71241
 ID AAN71241 standard; DNA; 1500 BP.

AC AAN71241;
 DT 02-MAY-1991 (first entry)
 DE Subtilisin gene from *Bacillus subtilis*.
 KW subtilisin; extracellular protease; amylase production; ss.
 OS *Bacillus subtilis*.

XX Key Location/Qualifiers
 FH sig_peptide 137..205
 FT /*tag= a
 FT mat_peptide 455..1279
 FT /*tag= b
 FT misc_RNA 206..454
 FT /*tag= c
 FT /label= pro sequence

XX EP246678-A.

XX 25-NOV-1987.

XX 01-JAN-1987; 87EP-0200590.

XX 01-JAN-1987; 87EP-0200590.

XX 24-JUN-1983; 83US-0507419.

XX 29-MAY-1984; 84US-0614491.

XX 29-MAY-1984; 84US-0614612.

XX 29-MAY-1984; 84US-0614615.

XX 29-MAY-1984; 84US-0614616.

XX 29-MAY-1984; 84US-0614617.

XX (GETH) GENENTECH INC.

XX Bott RR, Ferrari E, Wells JA, Estell DA, Henner DJ;

XX WPI: 1987-328920/47.

XX P-PSDB; AAP71060.

XX *Bacillus* strains not excreting subtilisin or neutral protease -

XX obtd. by recombinant DNA procedures, useful for enzyme prodn.

XX esp. of hydrolase(s) such as amylase

XX Example; Fig 7; 71pp; English.

XX B.subtilis 1168 chromosomal DNA was digested with EcoRI. A single
 CC 6kb fragment hybridised to a fragment from the C-terminus of the
 CC subtilisin structural gene in pS4 (see AAN71240). It was ligated to
 CC EcoRI-cut pBS42 and the ligation mixture used to transform *E.coli*
 CC ATCC 31446. Plasmid DNA was prepared from a pooled suspension of
 CC transformant colonies and used to transform a protease deficient
 CC strain of *B.subtilis* (B684). Plasmid DNA from protease producing
 CC colonies was digested with EcoRI and examined by Southern blot
 CC analysis to isolate the 6kb fragment. A positive clone was identified
 CC containing a plasmid designated pS168.1. Three HincII fragments
 CC and a HincIII-EcoRI fragment were ligated into M13 vectors and
 CC sequenced to obtain the entire subtilisin sequence.

SQ Sequence 1500 BP; 443 A; 343 C; 328 G; 386 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.39e-14 Length: 1500
 Score: 338.50 Matches: 133
 Percent Similarity: 40.58% Conservative: 76
 Best Local Similarity: 25.83% Mismatches: 126
 Query Match: 6.03% Indels: 126
 DB: 8 Gaps: 21
 US-09-830-837-6 (1-1052) x AAN71241 (1-1500)
 QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
 Db 233 AGCAGTACA-----GAAAAGAAATACATTGCGGATT-----AAA 268
 QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAlaAspAsnTrpArgIle 87
 Db 269 CAGACAATGAGTGCATGATGTTCCGCCAAGAAAAGGATGTTATT----- 313
 QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
 Db 313 ----- 313
 QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
 Db 314 ---TCGTGAAAAGCGGAAAGGTTCAAAGCAATTTAAGTATGTTAAACGCGCGCAGCA 370
 QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
 Db 371 ACATTGGATGAAAAGCTGTAAGAAATTGAAA-----AAAGATCCGAGCGTTGCA 421
 QY 147 CysAsnGluThrArgTyrSerGlnLysTyrGlnSerSerArgProLeuArgAlaSer 166
 Db 421 ----- 421
 QY 167 LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeu 186
 Db 422 -----TATGTGGAAGAAGATCATATTGCACATGAATATGCG 457
 QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
 Db 458 CAATCTGTTCTTATGCGCATTTCTCAA---ATTAAAGCGCGCGCTCTTCACTCTCAAGGC 514
 QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
 Db 515 TACACAGGCTTAACGTAAAGTAGTGTATGACAGCGGAGATTGACCTTCTCATCTCT 574
 QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
 Db 575 GACTTA---AACGTCAGAGCGGAGCAAGCTTCGTACCTTCTCAAAACAAACCCATACCAG 631
 QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
 Db 632 GACGCGAGTCTCACGCTACGCGATGAGCGGTACGATTCGCCCTCTTAATACTCAATC 691
 QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheTrpAsnAsn 281
 Db 692 GGTGTTCTGGCGTTAGCCCCACGCCATCATATTATGACAGTAAGTGTGATTCACAA 751
 QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTrpAlaIleLeuLysLysIle 301
 Db 752 GGAAGCGGCAATATAGTGGATTATTAACGGCATTTGAGTGGCCATTTCCACAATATG 811
 QY 302 AspValLeuAsnLeuSerIleGlyProAspPheMetasp-----HisProPheVal 319
 Db 812 GATGTTATCAACATGATGCGTTCGCGACCTACTGTTCTACAGCGCTGAAAACAGTCGTT 871
 QY 320 AspLysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAsp 339
 Db 872 GACAAAGCC-----GTTTCCAGCGGTATGCTGCTGCTCCGCGCGGAAACGAA 922
 QY 340 GlyPro-----LeuTyrGlyThrLeuAsnProAlaAspGlnMetAspValIleGly 357

CC subtilisins can be used as the second part of the fusion protein. The
 CC mutated versions of subtilisin are incapable of autolysis. The
 CC maturation. The DNA fusion sequences are useful for the production of
 CC heterologous proteins. As the subtilisin mutants are incapable of
 CC autolysis, the mutants are therefore bound to the cell membrane
 CC membrane. The inability of the mutant protein to leave the cell membrane
 CC means that they are not able to degrade desired proteins when
 CC heterologous protein production is being carried out in Bacillus cells.
 CC Heterologous proteins are sequestered at cell membranes and can be easily
 CC isolated and released by the action of enzymatically active subtilisin.
 XX
 SQ Sequence 1500 BP; 443 A; 343 C; 328 G; 386 T; 0 other;

Alignment Scores:

Pred. No.: 1,39e-14 Length: 1500
 Score: 338.50 Matches: 133
 Percent Similarity: 40.58% Conservative: 76
 Best Local Similarity: 25.83% Mismatches: 180
 Query Match: 6.03% Indels: 126
 DB: 20 Gaps: 21

US-09-830-837-6 (1-1052) x AA206712 (1-1500)

QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
 DB 233 AGCGTACA-----GAAAGAAATACATTCTCGATT-----AAA 268
 QY 68 AlaArgAsnSerPheIleSerAlaLeuLysSerSerGluValAspAsnTrpArgIle 87
 DB 269 CAGACAAATGAGTGCATGAGTTCGCCCAAGAAAGAGATTTATT----- 313
 QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
 DB 313 ----- 313
 QY 108 GluLysGlnLysAlaGluLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
 DB 314 ---TCTGAAAAGCGGAAAGGTTCAAAAGCAATTAAGTATGTTAAGCGCGCGCAGCA 370
 QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
 DB 371 ACATTGGATGAAAAGCTGTAAGAAGATTGAAA-----AAAGATCCGAGCGTTGCA 421
 QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer 166
 DB 421 ----- 421
 QY 167 LeuSerLeuGlySerGlyPheThrPheHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
 DB 422 -----TATCTGGAAGACATCATTTGCACATGAATATGCG 457
 QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
 DB 458 CAATCTGTCTTATGCGATTTCTCAA---ATTAAAGCGCGGCTCTTCACTCTCAAGGC 514
 QY 207 TyrThrGlyAlaAsnValArgValAlaPheAspThrGlyLeuSerGluLysHisPro 226
 DB 515 TACACAGGCTTAACGTAAAGTAGCTGTATCGACAGCGGAATTGACTCTTCTCATCT 574
 QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
 DB 575 GACTTA---AACGTACAGCGCGGAGCAAGCTTCGTACCTTCTGAAACAAACCCATACCAG 631
 QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
 DB 632 GACGGCAGTTCTCAGGTAGCATGTAGCGGTACGATTGCGGCTCTTAACTCAATC 691
 QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
 DB 692 GGTGTTCTGGCGTTAGCCCAAGCGCATCATATATGACAGTAAAGTCTGATTCACAA 751
 QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIle 301
 DB 301 ----- 301

DB 752 GGAAGCGGCCAATATACGTGGATTATTAAACGGCATTTGAGTGGCCATTTCCAAACAATATG 811
 QY 302 AspValLeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheVal 319
 DB 812 GATGTTATCAACATGAGCTTGGCGACCTACTGTTCTACACGCTGAAAACAGTCGCT 871
 QY 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
 DB 872 GACAAAGCC-----GTTTCCAGCGTATCGTCTGCTGCCGACGCGGAAACGAA 922
 QY 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
 DB 923 GGTTCATCCGGAAGCACAACAGCAGTCCGCTACCTCAAAATATCTTCTATTATTCGA 982
 QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377
 DB 983 GTAGTGGCGGTAACACAGCAGCAACAAAGAGCTTCATCTCCAGCGCAGTTCT----- 1036
 QY 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrGlyAlaGly 397
 DB 1037 ---GAGCTT-----GATGTGATGCTCTCGCGTGTC 1066
 QY 398 ValArgGlySerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
 DB 1067 ATCCAAAGCAGCTTCTGGAGCAGCTTACGCGCTTATACCGAAGCTTCAAGCGACT 1126
 QY 418 ProValAlaGlyAlaValThrLeuValSerThrValGlnLysArgGluLeuVal 437
 DB 1127 CCTCAGCTTGGCGGAGCAGCGTAAATCTTCT-----AAGCAGCCGACTTGG 1177
 QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsn 457
 DB 1178 ACAACGCGCAAGTCCGTGATGTTAGAAAGCAGCTCAACATATCTTGA-----AAC 1231
 QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAla-TyrGlnIleLeuAsnSe 477
 DB 1232 TCTTCTTACTATGAAAGGTTAATCAACGCTACAAGCAGCTCACAAT-----AATAG 1285
 QY 477 TyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMe 497
 DB 1286 TAAAGAAGCAGCTT-----CCTCCATACCTGCTCTTTTATTGTCAGCAT-- 1334
 QY 497 ttrpTyrCysSerGlnProIleTyr-TyrGlyMetProThrValValAsnValTh 517
 DB 1335 -----CCTGATGTTCCGGCGCATTC 1354
 QY 517 rIleLeuAsnGlyMet-----GlyValThrGlyArgI 528
 DB 1355 TCTTTTCTCCGATGTTGAATCCGTTCCATGATCGACGATGGCTGCTCTGAAATC 1414
 QY 528 eValAspLysProAspTrpGlnProTyrIleProGlnAsnGly 542
 DB 1415 TTCACAAGCAGCGGAGGATCAACCTGCTCAGCGCCGTCACGCC 1457
 RESULT 22
 AAQ03536
 ID AAQ03536 standard; DNA; 1499 BP.
 XX
 AC AAQ03536;
 XX
 DT 16-AUG-1990 (first entry)
 XX
 DE Subtilisin gene.
 XX
 KW Carbonyl hydrolase; subtilisin; neutral protease; ds.
 XX
 OS Bacillus subtilis.
 XX
 FH Key Location/Qualifiers
 FT CDS 137..1280
 FT /*tag= a
 PN EP357157-A.

XX 07-MAR-1990.
 XX PD
 XX PF
 XX 22-JUN-1984; 84EP-0202584.
 XX PR
 XX 24-JUN-1983; 83US-0507419.
 XX PR
 XX 29-MAY-1984; 84US-0614612.
 XX PR
 XX 29-MAY-1984; 84US-0614615.
 XX PR
 XX 29-MAY-1984; 84US-0614616.
 XX PR
 XX 29-MAY-1984; 84US-0614617.
 XX PR
 XX 29-MAY-1984; 84US-0614491.
 XX PA (GETH) GENENTECH INC.
 XX PI
 XX Bött RR, Estell DA, Ferrari E, Henner DJ, Wells JA;
 XX WPI; 1990-068909/10.
 XX DR P-PSDB; AAR03737.
 XX PT Mutant prokaryotic carbonyl hydrolase enzymes -
 PT obtd. by site-directed oligo-nucleotide mutagenesis, used in
 PT food processing and cleaning industries.
 XX PS Claim 16; Fig 7; 39pp; English.
 XX CC Probe derived from subtilisin gene was used to isolate carboxyl
 CC hydrolase gene, mutant versions of which exhibit different oxidative
 CC stability and/or pH activity.
 XX SQ Sequence 1499 BP; 443 A; 342 C; 328 G; 386 T; 0 other;

Alignment Scores:

Pred. No.: 1,51e-14 Length: 1499
 Score: 338.00 Matches: 124
 Percent Similarity: 41.96% Conservative: 69
 Best Local Similarity: 26.96% Mismatches: 164
 Query Match: 6.02% Indels: 103
 DB: 11 Gaps: 19

US-09-830-837-6 (1-1052) x AAQ03536 (1-1499)

QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
 DB 233 AGCAGTACA-----GAAAAGAAATACATGTCGGATTT-----AAA 268
 QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIle 87
 DB 269 CAGACAAATGAGTCCCATGATGTCGCCAAGAAAGAGATGTTATTT----- 313
 QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
 DB 313 ----- 313
 QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
 DB 314 ---TCTGAAAGCGGAAAGGTTCAAAAGCAATTTAAGTATGTTAACCGCGCGCAGCA 370
 QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
 DB 371 ACATGGATGAAAGAGCTTAAGAAATTGAA-----AAGATCCGAGCGTGTGA 421
 QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer 166
 DB 421 ----- 421
 QY 167 LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeu 186
 DB 422 -----TATGTGGAAGAGATCATATTGCACATGAATATGCG 457
 QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
 DB 458 CAATCTGTTCTTATGCGATTCTCAAA---ATTAAAGCGCGGCTCTTCACTCTCAAGGC 514

QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
 DB 515 TACACAGGCTCTAAACGTAAAGTAGCTGTTATGCACAGCGGAATGACTCTTCTCATCCT 574
 QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
 DB 575 GACTTA---AACGTGAGAGCGGAGCAAGCTTCGTACTCTCTGAAACAACCCATACCAG 631
 QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
 DB 632 GACGGCAGCTTCTCAGCGTACGCATGTAGCGGTACGATTCGCGCTCTTAATACTCAATC 691
 QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
 DB 692 GGTGTTCTGGCGCTTAGCCCAAGCGCATCATTTATATGCAGTAAAGGCTGTGATTCACA 751
 QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIle 301
 DB 752 GGAAGCGGCAATATAGCTGGATTATTAACGGCATTGAGTGGGCCATTTCCAAACAATATG 811
 QY 302 AspValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheVal 319
 DB 812 GATGTTATCAACATGAGCTTGGCGACCTACTGGTCTTACACGCGTGAACACAGTCGTT 871
 QY 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
 DB 872 GACAAAGCC-----GTTTCCAGCGGTATCGTCTGTCGCGCAGCGCGGAAGAA 922
 QY 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
 DB 923 GGTTCATCCGGAAGCACAGCAGTCGGCTACCTCGCAAAATATCTTCTACTATTGCA 982
 QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377
 DB 983 GTAGTGGCGGTAAACAGCAGCACCAAGAGCTTCATCTCCAGCGCAGGTTCT----- 1036
 QY 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly 397
 DB 1037 ---GAGCTT-----GATGTGATGGCTCCTCGCGTGTCC 1066
 QY 398 ValArgLysSerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
 DB 1067 ATCCAAGCACACTTCTCGAGCCTTACGGCGCTTATACGGAAGCTCATCGCAGTGGCGACT 1126
 QY 418 ProValValAlaGlyAlaValThrLeuValSerThrValGlnLysArgGluLeuVal 437
 DB 1127 CCTCAGTGGCGGACGACGAGCGTTAATCTTCT-----AAGCACCGCAGCTGG 1177
 QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn 457
 DB 1178 ACAACGCGCAAGTCCGTGATCGTTTAGAAAGCACTGCAACATATCTTGA-----AAC 1231
 QY 458 MetPheGlnGlyHisGlyLysLeuAspLeuLeuArgAla-TyrGlnIleLeuAsnSe 477
 DB 1232 TCTTTCTACTATGGAAGGGTTAATCAACGTACAAGCAGCTGCACAA-----AATAG 1285
 QY 477 rTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyr 496
 DB 1286 TAAAGAAGACGAGTT-----CCTCCATACCTGCTCTTTTATTATTGTGCAGCAT 1334

RESULT 23

AAAX25097/C

ID AAX25097 standard; DNA; 2760 BP.

XX AC AAX25097;

XX DT 05-JUL-1999 (first entry)

XX DE Bacillus subtilis metalloprotease YHEN DNA.

XX KW Metalloprotease; protease; YHEN; textile; detergent; feedstuff;

XX KW animal feed; host expression system; ss.

XX

QY 477 rTyrLysProGlnAlaSerProSerTyrIleAspLeuThrGluCysProTyr 496
 Db 1641 TAAAGAAGCAGGTT-----CCTCATACCTGCTCTTTTATTGTGTCAGCAT 1593
 RESULT 24
 AAD19518
 ID AAD19518 standard; DNA; 1180 BP.
 AC AAD19518;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Subtilisin E DNA, used to generate single-stranded DNA template.
 XX
 XX Template-mediated recombination; nucleic acid isolation; subtilisin E;
 KW therapeutic protein; vaccine; ds.
 XX
 OS Unidentified.
 XX
 FH Key
 FT misc_feature 1..16
 FT Location/Qualifiers
 FT /tag= a
 FT /note= "Sequences that optionally do not correspond
 FT to wild-type sequences which may be replaced by upstream
 FT regulatory regions and vector supplied sequences
 FT depending on the cloning system in use"
 FT 1163..1180
 FT misc_feature
 FT /tag= b
 FT /note= "Sequences that optionally do not correspond
 FT to wild-type sequences which may be replaced by upstream
 FT regulatory regions and vector supplied sequences
 FT depending on the cloning system in use"
 FT 1180..1180
 XX
 PN WO200164864-A2.
 XX
 XX 07-SEP-2001.
 XX
 XX 28-FEB-2001; 2001WO-US06775.
 XX
 XX 28-FEB-2000; 2000US-185244P.
 XX 29-FEB-2000; 2000US-185815P.
 XX 01-MAR-2000; 2000US-186247P.
 XX 02-MAR-2000; 2000US-186482P.
 XX 06-SEP-2000; 2000US-0656549.
 XX
 PA (MAXY-) MAXYGEN INC.
 XX
 XX Affholter JA, Cox A, Ness JE, Carr B;
 XX WPI; 2001-616208/71.
 XX
 XX Generating chimeric nucleic acids to produce therapeutics comprises
 XX hybridizing nucleic acids and nicking and elongating regions that are
 XX non-hybridized .
 XX
 PS Disclosure; Fig 6; 172pp; English.
 XX
 XX The invention relates to a method of single-stranded nucleic acid
 XX template-mediated recombination and nucleic acid fragment isolation.
 XX Methods include polymerase and polymerase-free recombination of
 XX nucleic acid fragments to generate chimeric nucleic acid sequences.
 XX This method is useful for isolating nucleic acid fragments,
 XX combinatorially assembling nucleic acids and for producing chimeric
 XX nucleic acids. The method is useful for single stranded nucleic
 XX acid template mediated recombination and nucleic acid fragment
 XX isolation. Chimeric genes can be generated to produce therapeutic
 XX proteins with enhanced activity and vaccines can be produced. The
 XX present sequence is subtilisin E DNA, used to generate single-stranded
 XX DNA template.
 XX
 XX Sequence 1180 BP; 339 A; 262 C; 269 G; 276 T; 34 other;

Alignment Scores:
 Pred. No.: 1..2e-14 Length: 1180
 Score: 337.50 Matches: 117
 Percent Similarity: 41.97% Conservative: 66
 Best Local Similarity: 26.83% Mismatches: 156
 Query Match: 6.01% Indels: 97
 DB: 22 Gaps: 17
 US-09-830-837-6 (1-1052) x AAD19518 (1-1180)
 QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
 Db 113 AGCAGTACA-----GAAAGAATAATATGTCGGATT 148
 QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerGluValAlaValAsnTyrArgile 87
 Db 149 CAGCAATGAGTCCCATGAGTTCGCCCAAGAAAGATGTTATT----- 193
 QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
 Db 193 ----- 193
 QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
 Db 194 ---TCTCAAAAGCGCGGAAAGGTTCAAAAGCAATTTAAGTATGTTTAAACGCGCGCAGCA 250
 QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
 Db 251 ACATTGGATGAAAGAGCTGTAAGAAAGTGA---AAAGATCCGACGGTTCGA 301
 QY 147 CysAsnGluThrArgTyrSerGlnLysTyrSerArgProLeuArgAlaSer 166
 Db 301 ----- 301
 QY 167 LeuSerLeuGlySerGlyPheThrHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
 Db 302 -----TATGTGAAGAAGATCATATATGCAATGATATGCG 337
 QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTyrGlnMetGly 206
 Db 338 CAATCTGTCTTATGGCATTTCTCAA---ATTAAAGCGCGGCTCTTCACTCTCAAGGC 394
 QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
 Db 395 TACACAGGCTCTAACGTAAAGTAGCTGTATGACAGCGGAATGACTTCTCTCATCCT 454
 QY 227 HisPheLysAsnValLysGluArgThrAsnTyrThrAsnGluArgThr-----LeuAsp 244
 Db 455 GACTTA---AACGTGAGCGCGGAGCAAGCTTCTGCTTCTTCTGAAACAAACCATACAG 511
 QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
 Db 512 GACGGCAGTCTCAGCGTACGCATGTAGCGGTGACGATTCGCGCTCTTAATACTCAATC 571
 QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
 Db 572 GGTGTTCTGGCGGTAGCCCAAGCGCATCATATATGACAGTAAAGAGTCTGTGATCAACA 631
 QY 282 GlnValSerTyrThrSerTyrPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIle 301
 Db 632 GGAAGCGGCAATATATGATTAATTAACGGCATTCAGTGGGCCATTTCCAAACAATATG 691
 QY 302 AspValLeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheVal 319
 Db 692 GATGTTATCAACATGAGCCTTGGCGGACCTTACTGGTCTACAGCGCTGAAACACGTCGTT 751
 QY 320 AspLysValTyrGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
 Db 752 GACAAAGCC-----GTTTCCAGCGGTATCGTGTGCTGCCCGCAGCGGAAACGAA 802
 QY 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
 Db 803 GGTTCATCGGAAGCACAGCAGTGGCTACCTTACCTTCTGCAAAATATATCTTCTACTATTGCA 862


```
QY 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
   |||
   |||:
   |||:
Db 923 GGTTCATCCGGAACACACAGTCGGCTACCTGCAAAATATCTTCTACTATTGCA 982
   |||
   |||:
   |||:
QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377
   |||
   |||:
   |||:
Db 983 GTAGTCCGGTAAACACAGCAGCAACCAAGAGCTTCATCTCCAGCGAGTTCT----- 1036
   |||
   |||:
   |||:
QY 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly 397
   |||
   |||:
   |||:
Db 1037 ---GAGCTT-----GATGTGATGGCTCTCGGCTGCC 1066
   |||
   |||:
   |||:
QY 398 ValArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
   |||
   |||:
   |||:
Db 1067 ATCCAAAGCACACTTCCTGGAGCACTTACGGCGCTTATAACGGAAGCTCCATGGCGACT 1126
   |||
   |||:
   |||:
QY 418 ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
   |||
   |||:
   |||:
Db 1127 CCTCACGTTCCGGAGCAGCAGCGTTAATCTTTCT-----AAGCACCGCACTGG 1177
   |||
   |||:
   |||:
QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn 457
   |||
   |||:
   |||:
Db 1178 ACAAACCGCAAGTCCGTGATCGTTAGAAAGCACTGCAACATATCTTGA-----AAC 1231
   |||
   |||:
   |||:
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAla-TyrGlnIleLeuAsnSe 477
   |||
   |||:
   |||:
Db 1232 TCTTCTACTATGAAAGGGTTAATCAACGTACAGCAGCTGCACAAAT-----AATAG 1285
   |||
   |||:
   |||:
QY 477 rTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMe 497
   |||
   |||:
   |||:
Db 1286 TAAAGAAGACAGGTT-----CCTCATACCTGCTTCTTTTATTGTCAGCAT-- 1334
   |||
   |||:
   |||:
QY 497 tTrrProTyrCysSerGlnProIleTyrTyrGlyMetProThrValValAsnValTh 517
   |||
   |||:
   |||:
Db 1335 -----CCTGATGTTCCGGCGCATTC 1354
   |||
   |||:
   |||:
QY 517 rIleLeuAsnGlyMet-----GlyValThrGlyArgIle 528
   |||
   |||:
   |||:
Db 1355 TCTTCTTCTCCGCATGTTGAATCCGTTCCATGATCGACGGATGGCTGCTGAAATC 1414
   |||
   |||:
   |||:
QY 528 eValAspLysProAspTrpGlnProTyrLeuProGlnAsnGly 542
   |||
   |||:
   |||:
Db 1415 TTCACAAGACCGGAGGATCAACCTGCTCAGCCCCCGTCACGGC 1457
   |||
   |||:
   |||:
```

Search completed: June 2, 2003, 02:58:04
Job time : 775 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 2, 2003, 02:58:17 ; Search time 973 Seconds
(without alignments)
1457.842 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNLTWLLVLLCGKKH.....PRVKRPLMQVHPKTPSV 1052

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 845702 seqs, 674182571 residues

Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O/cgn2_1/USPTO.spool/US09830837/runat_23052003_181926_8115/app_query.fasta_1.1223
-DB=published.Applications_NA_QFMT-fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=25 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09830837 -CGN_1_1_201 @runat_23052003_181926_8115
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT
-DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
5617	100.0	4338	10	US-09-891-711-3	Sequence 3, Appli
5487	97.7	4198	10	US-09-891-711-5	Sequence 5, Appli
734	13.1	467	9	US-09-918-995-12598	Sequence 12598, A
393	7.0	300	10	US-09-960-352-8775	Sequence 8775, Ap

5	337.5	6.0	1074	10	US-09-837-235-14	Sequence 14, Appl
6	335.5	6.0	1140	10	US-09-920-118-13	Sequence 13, Appl
7	332.5	5.9	1146	10	US-09-920-118-15	Sequence 15, Appl
8	331	5.9	1497	7	US-08-322-678-6	Sequence 6, Appli
9	331	5.9	1497	7	US-10-033-325-1	Sequence 1, Appli
10	331	5.9	1497	10	US-09-060-854B-1	Sequence 1, Appli
11	320	5.7	1977	12	US-10-090-624-11	Sequence 11, Appl
12	302	5.4	1236	12	US-10-090-624-2	Sequence 2, Appli
13	302	5.4	1962	12	US-10-090-624-15	Sequence 15, Appl
14	285	5.1	4765	12	US-10-090-624-5	Sequence 5, Appli
15	282	5.0	1140	7	US-08-322-678-11	Sequence 11, Appl
16	279	5.0	1140	7	US-08-322-678-12	Sequence 12, Appl
17	261.5	4.7	576	10	US-09-920-118-21	Sequence 21, Appl
18	260.5	4.6	3788	10	US-09-927-827-33	Sequence 33, Appl
19	248.5	4.4	1485	10	US-09-974-300-1938	Sequence 1938, Ap
20	247	4.4	497	10	US-09-783-590-349	Sequence 349, Ap
21	241.5	4.3	1329	10	US-09-974-300-1934	Sequence 1934, Ap
22	229.5	4.1	1330	10	US-09-966-921A-5	Sequence 5, Appli
23	228	4.1	1306	10	US-09-966-921A-1	Sequence 1, Appli
24	227.5	4.1	840	9	US-10-209-812-1	Sequence 1, Appli
25	225.5	4.0	1397	10	US-09-974-300-1935	Sequence 1935, Ap
26	225	4.0	1971	10	US-09-974-300-1935	Sequence 1935, Ap
27	221	3.9	522	9	US-09-824-893A-71	Sequence 71, Appl
28	218	3.9	429	9	US-09-918-995-35748	Sequence 35748, A
29	217	3.9	522	9	US-09-824-893A-94	Sequence 94, Appl
30	217	3.9	3650	9	US-09-927-827-31	Sequence 31, Appl
31	216	3.8	522	9	US-09-824-893A-55	Sequence 55, Appl
32	215	3.8	522	9	US-09-824-893A-112	Sequence 112, App
33	212.5	3.8	3825	10	US-09-974-300-1925	Sequence 1925, Ap
34	212	3.8	522	9	US-09-824-893A-107	Sequence 107, App
35	211	3.8	522	9	US-09-824-893A-97	Sequence 97, Appl
36	211	3.8	522	9	US-09-824-893A-128	Sequence 128, App
37	210	3.7	522	9	US-09-824-893A-57	Sequence 57, Appl
38	208	3.7	522	9	US-09-824-893A-45	Sequence 45, Appl
39	208	3.7	522	9	US-09-824-893A-102	Sequence 102, App
40	207	3.7	522	9	US-09-824-893A-84	Sequence 84, Appl
41	206	3.7	522	9	US-09-824-893A-74	Sequence 74, Appl
42	206	3.7	522	9	US-09-824-893A-104	Sequence 104, App
43	206	3.7	522	9	US-09-824-893A-110	Sequence 110, App
44	205	3.6	522	9	US-09-824-893A-123	Sequence 123, App
45	204	3.6	522	9	US-09-824-893A-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1

US-09-891-711-3
; Sequence 3, Application US/09891711
; Patent No. US20020082404A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26 *to date*
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (497)..(3655)
US-09-891-711-3

Alignment Scores:			
Pred. No.:	0	Length:	4338
Score:	5617.00	Matches:	1052
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

[illegible]


```
QY 721 ValIlePheSerAspTyrTrpAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
Db 2657 GTCATCTTCAGTGACTGGTACACACTTCTGTATGAGAAAGTGAAGTTTATGATGAA 2716
QY 741 AsnThrArgGlnTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGlu 760
Db 2717 AACACAAAGCAGTGGTGGATGCGGATACCGGAGGAGCTAACATCCAGCTCTGAATGAG 2776
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGluPheThrLeu 780
Db 2777 CTGCTCTCTGTGGGAACATGGGGTTACGGATGGCTGTATGAAGGGAGTTACCCCTG 2836
QY 781 AlaAsnHisAspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
Db 2837 GCCAACCATGACATGATTATGTCGTGAGGTGCAGCATCGCAAGTTTCCAGAAGATGCG 2896
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnIleThrAla 820
Db 2897 GTCGTGATACACAGACTTTTCAAGGACCAAGGATTGGAGGTTTAAAGCAGGAACAGCA 2956
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
Db 2957 GTTGTGAAACCTCCCACTTTTGGGACTTTATCAGATTCCAGCTCAGGCTGGAGGCCG 3016
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspSerHisArgGlnLysAspCysPhe 860
Db 3017 ATTGTACTGTATGGGACTCCAATTCTGTGGATGACAGTCACCGAAGGACTGCTTT 3076
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
Db 3077 TGGCTTCTGGATGCCCTCTCCAGTACATCGTATGGGTGCACACGCCCTAGCCTCAGT 3136
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
Db 3137 CACTCTGGGAACCGCCAGCGCCCTCCAGTGAGCAGGCTCAGTCACTCCAGAGGATG 3196
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
Db 3197 GAAGGAACCATCTTCATCGGTACTCCAAGTTCTGGAGCCCATTTGGAGACCCCAAA 3256
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
Db 3257 CCTCGCCCTCTACACGCTGTCCACGCTTGTCTTGGGCCAAGCCACAGCCTTTAAACGAG 3316
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
Db 3317 ACGGCCCCAGTAACCTTTGGAAACATCAGAGCTACTCTCCATTGACCTGGACACAGGTG 3376
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3377 GTGTTACCAACTTTCGATCGAATCGCCCTCAAGTGAGGCCCTTGTCCCTCGAGAGAGC 3436
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnValGlyGln 1000
Db 3437 GCGCCCTGGGACATCTCCTGGAGGATCATGCTGCGCCGTACAAACAGGAGGTGGCCAG 3496
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValIleAlaPheValValGln 1020
Db 3497 ACATTCTCTCTTCCCTTCGGGAGCATGGTGGCTCGGCCCTCTTGTGGTACAA 3556
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3557 ATCAACAGGCCAAGACGAGCCGAGGAGGAGGAGGCCAGGCTGAGCGCCCGCAGCTC 3616
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3617 ATGCACAGGTTCCACCGCCAAAGACCCCTTCGGTG 3652

RESULT 2
US-09-891-711-5
; Sequence 5, Application US/09891711
; Patent No. US20020082404A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4198
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (387)..(3545)
US-09-891-711-5

Alignment Scores:
Pred. No.: 0 Length: 4198
Score: 5487.00 Matches: 1021
Percent Similarity: 98.57% Conservative: 16
Best Local Similarity: 97.05% Mismatches: 15
Query Match: 97.69% Indels: 0
DB: 10 Gaps: 0

US-09-830-837-6 (1-1052) x US-09-891-711-5 (1-4198)
QY 1 MetLysLeuValIleAsnIleTrpLeuLeuValValLeuLeuCysGlyLysLysHis 20
Db 387 ATGAAGCTCATCAACATCTGGCTTCTTCTGCTGGTGGTTTCTGTGTGGAAGAAGCAT 446
QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40
Db 447 CTGGGTGACAGGCTGGGGAAGAAGCGTTTGAAGGCATCATGCCCTAGCTGTCCAC 506
QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrIleValAlaPhe 60
Db 507 CTGACTTTGAAGGTGGAATTTCTCTCAACTGTGGTGAATATGAATATATTTGGCTTTC 566
QY 61 AsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSer 80
Db 567 AACGGATACTTCACGCCAAAGCTAGAAGCTATTTATTTCAAGTGTCTGTGAAGAGCAT 626
QY 81 GluValAspAsnTrpArgIleIleProArgAsnAsnProSerSerAspTyrProSerAsp 100
Db 627 GAAGTAGACAACCTGGAGAAATATACCTCGGAACAACCCATCCAGTACCTAGTATGAT 686
QY 101 PheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis 120
Db 687 TTTGAGGTGATTTCAGATAAAAGAGAAGCAGAGCGCGGCTGCTCACACTTGAAGATCAT 746
QY 121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu 140
Db 747 CCAAAATCAAGCGGGTGACACCTCAACGCAAGTCTTTCGTTCCTTGAAGTTTGCAGAA 806
QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg 160
Db 807 TCTGACCCCATTTGCGCATGTATGAACCTCGTGGAGCCAGAGTGGCAGTCAACGA 866
QY 161 ProLeuArgAlaSerLeuSerLeuGlySerGlyPheThrPheIleAlaThrGlyArgHis 180
Db 867 CCCTCTGAGAAGAGCCAGTCTCTCCCTGGGCTCTGGATTTCTGGCATGCAACAGGAACAT 926
QY 181 SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
Db 927 TCAAGCGCGGATTTGTGAGAGCCATTCCTCGACAGGTTTCCCGACAGATTCGAGGAGAT 986
QY 201 ValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGly 220
Db 987 GTCTGTGGCAGATGGGATACAGGTGCTAAATCTCAGGCTTCTGTTTGTACTGGG 1046
QY 221 LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240
```

1047 CTCAGTGAAGAATCCACACTTCAAGAAATGTGAAGGAGAGAAACCAACTGGACCAATGAG 1106
Db
241 ArgThrLeuAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
QY
1107 CGGACCTGGATGGCTGGCCATGGACATTTGTCAGGTGTGATTCGCCAGCATG 1166
Db
261 ArgGluCysGlnClyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
QY
1167 AGGAGTGCACGGGATTTGCCACAGATGCAGAGCTGCACATCTTCGGGTCTTTACCAAC 1226
Db
281 AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLys 300
QY
1227 AATCAGGTGCTTACACATCTTGTTTGGACGCTTTCACACTATGCCATCTCTAAGAG 1286
Db
301 IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisProPheValAsp 320
QY
1287 ATTGATGTTCTAAACCTTAGCATCGCGGCCCTGACTTCATGGATCATCCCTTTGTGC 1346
Db
321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
QY
1347 AAGGTGTGGGAATTAAACAGCTAAACATGTAATCATGGTTCTGCTATCGCAATGATGA 1406
Db
341 ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 360
QY
1407 CCTCTTATGGCACTCTGAATAACCCAGCTGATCAGATGGATGATGATGGAGTGGTGC 1466
Db
361 IleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrThrTrpGluLeu 380
QY
1467 ATTGACTTTGAAGATAACATCGCCGCTTTCTTCCAGGGAATGACTACCTGGGAATA 1526
Db
381 ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly 400
QY
1527 CCAGGAGCTATGTCGCTGGAACCTGACATTTGTCACCTATGTCGCGAGTGGCGGT 1586
Db
401 SerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420
QY
1587 TCGGTGTGAAGGGGGCTGCGGGCACTCTCAGGACCACTGCTGCTTCCCGAGTGGT 1646
Db
421 AlaGlyAlaValThrLeuValSerThrValGlnLysArgGluLeuValAsnProAla 440
QY
1647 GCTGGGCTGTACCTTGTAGTACGACATGTCAGAGGGAGCTAGTGAATCTGCC 1706
Db
441 SerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGlu 460
QY
1707 AGTGTGAGCAGCCCTGATTCATCAGCCCGGAGCTTCTGTGTGTACATGTTCCAG 1766
Db
461 GlnGlyHisGlyLysLeuAspLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480
QY
1767 CAAGGCCATGGCAAGCTGGATCTGCTGGAGCCTATCAGATCCTCAGCACTACAACCA 1826
Db
481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
QY
1827 CAGCGAGCTTGAGTCTAGCTACATCAGCTGAGTGTGCTACATGTGGCCTTAC 1886
Db
501 CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn 520
QY
1887 TGTCTTCAGCCCATCTACTATGAGGAATGCCAACAAATGTTAATGTACCACTCTCAAT 1946
Db
521 GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln 540
QY
1947 GGCATGGAGTCACAGGAAGATTTGGATTAAGCTGAGTGGCGGCCCTATTACACAG 2006
Db
541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerValIleuTyrProTrpSerGly 560
QY
2007 AATGAGACACATTAAGTGGCTTCTCTACTCTCAGTGTATGGCTTGGTCCAGC 2066
Db
561 TyrLeuAlaIleSerIleSerValThrLysAlaAlaSerTrpGluGlyIleAlaGln 580
QY
2067 TACCTGGCCATCTCATTTCTGTGACCAAGAGGAGCTTCTTGGGAAGGCTTGCACAG 2126
Db
581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu 600
QY
2127 GGTACATCATGATCAGGTGGCTTCCCGACGAGACGGAAGCAAAAATGTTGGTCCGAG 2186
Db
601 GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProAspSer 620
QY
2187 CATACTTCACAGTAGTGAAGCTTCCCATTAAGTGAAGATCATTTCCACCCCTCTCCGAGC 2246
Db
621 LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProAsp 640
QY
2247 AAGAGAGTCTCTGGGACAGTATCAACAACCTCCGCTACCCCCAGGCTACTTTCACAGG 2306
Db
641 AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn 660
QY
2307 GACAACCTGGATGAAGATGATCTTTAGACTGAATGGCGACATGTGCCACCAAT 2366
Db
661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
QY
2367 TTCAGGACATGTACCAGCACCTGCGCAGCATGGCTACTCTGCTGGAGGTGCTCGGTGC 2426
Db
681 ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
QY
2427 CCATTCAGTGTCTTGATGCTACACATGATGGCACTTTGCTCATGGTGAATGAAGAA 2486
Db
701 GlnTyrPheProGluLeuIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu 720
QY
2487 GAGTACTTCCAGGAGATTCACCAAGCTGAGGAGGAGCTGGCAATGGCTTTCCTC 2546
Db
721 ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
QY
2547 GTCACTTTCAGTCTGCTGACACATCTCTGTTATGAGAAAGTGAAGTGTACGATGAA 2606
Db
741 AsnThrArgGlnTrpTrpMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu 760
QY
2607 AACACAGGAGGTGGTGGATGCCAGATCTGAGGAGGAGCAACATCCAGCTCTGACAGAG 2666
Db
761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyLysPheThrLeu 780
QY
2667 CTGCTGTCTGTGGAACTGGGTTCAGCATGGCTTATGAAGGGAGTGTCCAGAGATGGT 2726
Db
781 AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
QY
2727 GCGAATCATGACATGATTTATGATCGGGATGAGCATGCCAAGTCTTCCAGAGATGGT 2786
Db
801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
QY
2787 GTTGTGATCAGACAGATTTCAAGGACCAAGGATGGAGGTCTTAAAAACAAGACAGCA 2846
Db
821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
QY
2847 GTTGTGAAATGTTCCCATTTTGGGCTTTATCAGATTCAGCTCAAGGTGGGCGCG 2906
Db
841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
QY
2907 ATCTGTGTGTATGGAGATTCATTTGCTTGGATGACAGTCCACAGACAGAGATGCTTT 2966
Db
861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
QY
2967 TGGCTTCTGGATGACCTCTTTCAGTACATCATATGCGTGAACCTCCAGGCTCCAGC 3026
Db
881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
QY
3027 CATTCAGGAACCGGACGCGCCACCCAGTGGAGTGGTGTGGCCCCCTCTCTGAAGAGTG 3086
Db
901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuLysAspProLys 920
QY
3087 GAAGGAAACCACTTTCATGATACTCCAAAGGTCTTGGAGCCCATCTGGAGACCCAAAA 3146
Db
921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
QY
3147 CCTGGCTCTTCCAGCCTGTCCACACTTGTCTATGGGCGCAAGCACAGCTTGTGAATGAG 3206
Db
941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
QY
3207 ACTGGCCCCAGTATCTTTGGAAACATCAGAGGTGCTCTCCATTCACCTGGACAAAGTA 3266
Db

QY 961 ValLeuProAspPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3267 GTGTTACCAACTTTGATCGAATCGCCCTCAAGTGAGACCTTTGTCCCTCGAGAAAGT 3326
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnValGlyGln 1000
Db 3327 GGTGCGCTGGGACATCTCTGGAGGATCATGCTGCGCGCTACAAACAGAGGTGGCCAG 3386
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValIleAlaPheValValGln 1020
Db 3387 ACCATCCCTGCTTTCCTCTCCGAGGCATGGTGGCCCTCTTTTGGGTACAG 3446
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3447 ATCAGCAAGCGCAAAAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3506
QY 1041 MetGlnGlnValHisProLysThrProSerVal 1052
Db 3507 ACACAGCAGACCCACCCACCAAGGACCCGCTCAGTG 3542

RESULT 3
US-09-918-995-12598
; Sequence 12598, Application US/09518995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12598
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(467)
; OTHER INFORMATION: n = A,T,C or G

US-09-918-995-12598

Alignment Scores:
Pred. No.: 1,93e-69 Length: 467
Score: 734.00 Matches: 136
Percent Similarity: 99.28% Conservative: 0
Best Local Similarity: 99.28% Mismatches: 1
Query Match: 13.07% Indels: 0
DB: 9 Gaps: 0

US-09-830-837-6 (1-1052) x US-09-918-995-12598 (1-467)

QY 710 LeuArgArgAspValAspAsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThr 729
Db 48 CTCGGAGGAGCGTGGACACGCGCTCTCGCTCATCTTCAGTGACTGGTATTACACT 107
QY 730 SerValMetArgLysValLysPheTyrAspGluAsnThrArgGlnTrpMetProAsp 749
Db 108 TCTGTTATGAGAAAGTGAAGTTTATGATGAAACACAAAGGAGGAGTGGTGGATCGCGGAT 167
QY 750 ThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPhe 769
Db 168 ACCGGAGGAGCTAACATCCAGCTCGAATGAGCTGCTGTGTGGAAACATGGGGTTC 227
QY 770 SerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTrpAlaSer 789
Db 228 AGCGATGGCTGTATGAGGGAGTTACCTCGCCACCAACATGATGATGATGATGATGATGATG 287
QY 790 GlyCysSerIleAlaLysPheProGluAspGlyValIleThrGlnThrPheLysAsp 809
|||||

Db 288 GGGTGCAGCATCCGCAAGTTTCCAGAAGATGGCGTGTGATAACACAGACTTTTCAAGGAC 347
QY 810 GlnGlyLeuGlnValLeuLysGlnGluThrAlaValValGluAsnValProIleLeuGly 829
Db 348 CAGGATTTGAGGTTTAAAGCAGGAACACAGAGTTGTTGAAACGTCCTCCATTTGGGA 407
QY 830 LeuTyrGlnIleProAlaGluGlyGlyArgIleValLeuTyrGlyAspSerAsn 848
Db 408 CTTTATCAGATCCAGCTGAGGCGCGGATTTGATGATGATGATGATGATGATGATGATGAT 464

RESULT 4
US-09-960-352-8775
; Sequence 8775, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8775
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 38-LIB188-011-01-E1-B6

US-09-960-352-8775

Alignment Scores:
Pred. No.: 3,7e-33 Length: 300
Score: 393.00 Matches: 70
Percent Similarity: 94.81% Conservative: 3
Best Local Similarity: 90.91% Mismatches: 4
Query Match: 7.00% Indels: 0
DB: 10 Gaps: 0

US-09-830-837-6 (1-1052) x US-09-960-352-8775 (1-300)

QY 832 GlnIleProAlaGluGlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAsp 851
Db 3 CAGCTCACAGCTGAGGCGCGCGCTGCTGCTGATGGAGACTCCAACTGCTTGAC 62
QY 852 AspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuGlnTrpThrSer 871
Db 63 GACAGTCACCGGCGAGAGGACTGCTTCTGGCTCCTGGATGCACTCCTGCACTGTCAGTTCACGTCG 122
QY 872 TyrGlyValThrProProSerLeuSerHisSerGlyAsnArgGlnArgProProSerGly 891
Db 123 TATGGGTGACG 182
QY 892 AlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHisArgTyr 908
Db 183 GCAGGCTCGCAGCG 233

RESULT 5
US-09-837-235-14
; Sequence 14, Application US/09837235
; Patent No. US20020061549A1
; GENERAL INFORMATION:
; APPLICANT: Marshall, Christopher
; APPLICANT: Hoffman, Alexander
; APPLICANT: Errico, Joseph
; APPLICANT: Marshall, Paul
; TITLE OF INVENTION: STABILIZED PROTEINS
; FILE REFERENCE: 9725-005-999
; CURRENT APPLICATION NUMBER: US/09/837,235
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US00/28595
; PRIOR FILING DATE: 2000-10-16

```
; PRIOR APPLICATION NUMBER: 60/159,763
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-837-235-14

Alignment Scores:
Pred. No.: 2,52e-26 Length: 1074
Score: 337.50 Matches: 117
Percent Similarity: 41.97% Conservative: 66
Best Local Similarity: 26.83% Mismatches: 156
Query Match: 6.01% Indels: 97
DB: 10 Gaps: 17

US-09-830-837-6 (1-1052) x US-09-837-235-14 (1-1074)
QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
DB 25 ACAGTACA-----GAAAGAATAATACATTGCGGATTT-----AAA 60
QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAlaAspAsnTrpArgIle 87
DB 61 CAGACAATGAGTCCATGAGTTCCGCCAAGAAAGAGATGTTATT----- 105
QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
DB 105 ----- 105
QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
DB 106 ---TCTGAAAAGCGCGAAGGTTCAAAGCAATTTAAGTATGTTAACCGCGCGCAGCA 162
QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
DB 163 ACATTGGATGAAAGCACTGTAAGAAATTCGAAA-----AAGATCCGACGGTTCGA 213
QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer 166
DB 213 ----- 213
QY 167 LeuSerLeuGlySerGlyPheThrPheHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
DB 214 -----TATGTGAAGAAGATCATATGTCACATGAATATCGG 249
QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
DB 250 CAATCTGTCTTATGCGCAATTCCTCAA---ATTAAAGCGCGGCTCTTCACTCTCAAGCC 306
QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
DB 307 TACACAGGCTCTAACGTAAGTAGCTGTTATCGACAGCGGAATGACTCTTCTCATCT 366
QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
DB 367 GACTTA---AACGTCAGAGCGCGAGCAAGCTGCTACCTCTGAAACAAACCCATACCAG 423
QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
DB 424 GAGCGAGTCTCAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483
QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
DB 484 GGTGTTCTGGCGTTAGCCCAAGCGCATCATATATAGTAAGTAAAGTCTGATTCACAA 543
QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIle 301
DB 544 GNAAGCGCCAAATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 603
QY 302 AspValLeuAsnLeuSerIleGlyCysProAspPheMetAsp-----HisProPheVal 319

; PRIOR APPLICATION NUMBER: 60/159,763
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-837-235-14

Alignment Scores:
Pred. No.: 4.53e-26 Length: 1140
Score: 335.50 Matches: 126
Percent Similarity: 39.18% Conservative: 66
Best Local Similarity: 25.71% Mismatches: 169
Query Match: 5.97% Indels: 129
DB: 10 Gaps: 17

US-09-830-837-6 (1-1052) x US-09-920-118-13 (1-1140)
QY 1 MetLysLeuValAsnIleTrpLeuLeuLeuValValLeuLeuCysGlyLysLysHis 20
DB 4 ATGAGAAAAGAGTTTGGCTGGATGCTGACGCGCTTAATG----- 48
QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIRECTED GENE ASSEMBLY
; APPLICANT: Menzel, Rolf
; FILE REFERENCE: 10424-003
; CURRENT APPLICATION NUMBER: US/09/920,118
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/222,134
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1140)
US-09-920-118-13

RESULT 6
US-09-920-118-13
; Sequence 13, Application US/09920118
; Patent No. US20020102734A1
; GENERAL INFORMATION:
; APPLICANT: Menzel, Rolf
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIRECTED GENE ASSEMBLY
; FILE REFERENCE: 10424-003
; CURRENT APPLICATION NUMBER: US/09/920,118
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/222,134
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1140)
US-09-920-118-13
```

```

Db 48 ----- 48
QY 41 LeuThrLeuLysValGluPheSerSerThrVal-----Val 52
Db 49 CTCGTGTTACAGTCCGCTTCAGCATTCGCGCTGCTCAGCGCGGCAAAATGTT 108
QY 53 GluTyrGluTyrIleValAlaPhe---AsnGlyTyrPheThrAlaLysAlaAsnSer 71
Db 109 GAAAGGATATATGTCGATTAAGTCGGAGTGAACACCCATCCCTCAAAAGGAC 168
QY 72 PheIleSerSerAlaLeuLysSerSerGluValAsp---AsnTrpArgIleIleProArg 90
Db 169 ATCATCAAGAGAGC-----GGCGGAAGAGTGCACAAGCAGTTAGAAATCATCAAGCG 222
QY 91 AsnAsnProSerSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGln 110
Db 223 GCAAAAGCGAAGCTAGAC----- 240
QY 111 LysAlaGlyLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArg 130
Db 241 AAGAAGCGCTTCAGGAAGTCAAAATGATCCGGATGTC----- 279
QY 131 LysValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThr 150
Db 280 -----GCTTATGTGAAGAGGAT----- 297
QY 151 ArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGly 170
Db 297 ----- 297
QY 171 SerGlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIlePro 190
Db 298 -----CAGTAGCTCATGCTTTGGCGCAACCGTTCCT 330
QY 191 ArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAla 210
Db 331 TAGCGGCAATT---CCTCTCATTAAGCGGACAAAGTGCAGCTCAAGGCTACAAGGGAGCG 387
QY 211 AsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsn 230
Db 388 AAGCTAAAGTCGCGCTCGGATACAGGAATCCAAAGCTTCTCATCCGGACTTGAACGTA 447
QY 231 ValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGly 250
Db 448 GTCGCGGAGCAAGCTTCGTAGCTGGCGAGCTTATAACACCGCGCAACGCGACGCGC 507
QY 251 ThrPheValAlaGlyValIleAlaSerMetArgGluCysGln-----GlyPheAla 267
Db 508 ACGCATGTTCCGCGTACAGTAGCTGGCTTGACAATAACAACGGGTGTATTAGCGGTGCG 567
QY 268 ProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSer 287
Db 568 CCGAAGCTATCTTGTACCGCGTTAAGTCTGAATTCAGCGGAGCGGATCTTACAGC 627
QY 288 TrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspValLeuAsnLeuSer 307
Db 628 GGCATTGTAAGCGGAATCGAGTGGCGACGACAAACGGCATGGATGTATTACATGAGC 687
QY 308 IleGlyGlyProAspPheMetAsp-----HisProPheValAspLysValTrpGluLeu 325
Db 688 CTTGGAGGACCATCAGGCTCAACAGCGATGAACAGCGGTTGACAATGCATAT----- 741
QY 326 ThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyProLeu-----Tyr 343
Db 742 ---GCAAGAGGGTGTGCTTGGCGGCTGCTGGGAACAGACAGCGGATCTTCAGGAACACG 798
QY 344 GlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyGlyIleAspPhe 363
Db 799 AATAACAATCGCTATCTCTGGAATACGACTCTGTATCGCATCTGCGCGGTAGACCT 858
QY 364 GluAspAsnIleAlaArgPheSerSerArgGlyMetThrThrTrpGluLeuProGlyGly 383
Db 1108 GGTCTGATCAATGTCGAAGCTCGCGCTCAA 1137

RESULT 7
US-09-920-118-15
; Sequence 15, Application US/09920118
; Patent No. US20020102734A1
; GENERAL INFORMATION:
; APPLICANT: Menzel, Rolf
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIRECTED GENE ASSEMBLY
; FILE REFERENCE: 10424-003
; CURRENT APPLICATION NUMBER: US/09/920,118
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/222,134
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 15
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
US-09-920-118-15
Alignment Scores:
Pred. No.: 9.59e-26 Length: 1146
Score: 332.50 Matches: 117
Percent Similarity: 41.74% Conservative: 65
Best Local Similarity: 26.83% Mismatches: 157
Query Match: 5.92% Indels: 97
DB: 10 Gaps: 17

US-09-830-837-6 (1-1052) x US-09-920-118-15 (1-1146)
QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
Db 97 AGCAGTACA-----GAAAGAATAACATTGTCGGATT-----AAA 132
QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIle 87
Db 133 CAGACAATGAGTCCCATGAGTTCGCCCAAGAAAAGATGTTATT----- 177
QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
Db 177 ----- 177
QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
Db 178 ---TCGGAAGAGGGGAGGAGTTCAACCAATTTAAGTATGTTAACGGCGCGGAGCA 234
QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
Db 1111 ----- 1111
```

235 ACATTGGTAAAGAGCTGTAAAGAAATGAAA-----AAAGATCGGCGCTTCA 285
QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer 166
Db 285 ----- 285
QY 167 LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
Db 286 -----TATGTGGAAGAAGATCATATGTCACATCAATATGCG 321
QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
Db 322 CAATCTGTCTTATGCAATTCCTCA-----ATTAAAGCGCGGCTCTTCACCTCAAGGC 378
QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
Db 379 TACACAGCTCTTACCTAAAGTAGCTGTTCGACAGCGAATGACCTTCTCTATCT 438
QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
Db 439 GACTTA---AACGTACAGAGCGGAGCAAGCTTCTGACCTTCTGAAACAACCCATACCAG 495
QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 496 GACGCGAGTCTTACCGGTACGATGTAGCGGTACGATGCGCTCTTATTAACCTCAATC 555
QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
Db 556 GGTGTCTTGGCGGTAGCGCAAGCGCATATTATATGCAAGTAAAGTGCCTGATTCACAA 615
QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysIle 301
Db 616 GGAAGCGGCCAATATAGCTGGATTATTAAAGCGCATTTGAGTGGCCATTTCCAAACATATG 675
QY 302 AspValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheVal 319
Db 676 GATGTATACATAGAGCTTGGCGGACCTACTGCTTCTACAGCGCTGAAACAGCTCTT 735
QY 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
Db 736 GACAAAGCC-----GTTTCCAGCGGTATGCTGTGCTGCGGCGGCGGAAACGAA 786
QY 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
Db 787 GGTTCATCCGAGCAACAGCACAGCTGGCTACCGCTGACCAAAATATCTCTTACTATGCA 846
QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377
Db 847 GTAGTGGCGTAAACAGCACAGCAACCAAGAGCTTCAATCTCCAGCGCAGGTCT----- 900
QY 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly 397
Db 901 ---GAGCTT-----GATGTATGCTCTCTCGCGGTGCC 930
QY 398 ValArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
Db 931 ATCCAAAGCACACTCTCTGGAGGCACTTACGGCGCTTATAACGGAACGCTCATGCGGCA 990
QY 418 ProValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
Db 991 CCTCAGTGGCCGAGCAGCAGCTTATCTTCT-----AAGCACCAGCTTGG 1041
QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn 457
Db 1042 ACAACGCGCAGTCCGTGATCGTTAGAACACACCTGCAACATATCTTGA-----AAC 1095
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuArgAlaTyrGln 473
Db 1096 TCTTCTACTATGGAAGAGGTATATCAACGTACAGCAGCTGCACAA 1143
RESULT 8
US-08-322-678-6
; Sequence 6, Application US/08322678

Publication No. US20030077807A1
GENERAL INFORMATION:
APPLICANT: Graycar, Thomas P
APPLICANT: Bott, Richard R
APPLICANT: Wilson, Lori J
TITLE OF INVENTION: Subtilisin Variants
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc
STREET: 180 Kimball Way
CITY: So. San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/322,678
FILING DATE: 13-Oct-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hord, Margaret A.
REGISTRATION NUMBER: 33,401
REFERENCE/DOCKET NUMBER: GC235-2
TELEPHONE: (415) 742-7536
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1497 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-322-678-6

Alignment Scores:
Pred. No.: 2,12e-25 Length: 1497
Score: 331.00 Matches: 124
Percent Similarity: 40.40% Conservative: 59
Best Local Similarity: 27.37% Mismatches: 146
Query Match: 5.89% Indels: 124
DB: 7 Gaps: 19

US-09-830-837-6 (1-1052) x US-08-322-678-6 (1-1497)
QY 8 LeuLeuLeuValValLeuLeuCysGlyLysHisLeuGlyAspArgLeuGluLys 27
Db 45 ATAATCTCTATTGGTTATCTCTCAATGAAAGAGGAGGAGGATAAGAGTGAAGGC 104
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 105 AAAAAAGATGATGATGATGCTTGTCTTGTCTTATGCG-----TTA 143
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 144 ATCTTTAGATGGCTTTCGCGACACATCTCTCCCGAGCGCGGAGGAAATCAACGGG 203
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 204 GAAAGAAATATATTGCGGGTT-----AACACACATATGACGACG 245
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn 92
Db 246 ATGACGCGCGCTAGAGAAGATGTCATT----- 275
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 276 -----TCTGAAAAGGC 287

```
QY 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArgLys 131
Db 288 GCGAAGTGTAAAAGCAATTCAAATATGTAGACGAGCTTCAGTCACATTAAACGAAAA 347
QY 132 ValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArg 151
Db 348 GCTGTAAAGAAATGTAAA-----AAAGACCGGAGCGTCCT----- 383
QY 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySer 171
Db 383 ----- 383
QY 172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArg 191
Db 384 -----TACGTTGAAGAAGATCATCGTACGACATCGGTACGCGCGTCGCTTAC 434
QY 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsn 211
Db 435 GCGGTATCAAA---ATTAAAGCCCTGCTCTGCATCTCAAGGCTACACTGGATCAAT 491
QY 212 ValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 492 GTTAAGTAGCGGTATCGACACGCGTATCGATCTCTCATCTCTGATTTAAGGTAGCA 551
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
Db 552 AGCGGAGCCAGCATGGTCTCTCTGAACAAAT-----CCTTTCCAAGAC 596
QY 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 597 ACAACTCTCAGGAAGTACGTTGCGGCGACAGTTCGCGCTCTTAATAACTCAATCGGT 656
QY 263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
Db 657 GTATTAGCGGTGCGCAAGCGCATCATTTAGCTGTAAAGCTTCTCGGTGTCAGCGT 716
QY 283 ValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAsp 302
Db 717 TCCGGCCAAATACAGCTGGATCATTAAGGAATCGAGTGGCGCATCGCAACAATATGGAC 776
QY 303 ValLeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheValAsp 320
Db 777 GTTATTAACTATGAGCTCGCGGACCTCTGGTTCGTGCTGCTTTTAAAGCGGAGTGTAT 836
QY 321 LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly 340
Db 837 AAAGCC-----GTTGCATCCGCGCTGCTAGTCTTCGCGCAGCGGTACGAAGC 887
QY 341 Pro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
Db 888 ACTTCGCGCAGCTCAAGACAGTGGGCTACCTGGTAAATACCTCTCTGTCATTGCAGTA 947
QY 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrp 378
Db 948 GCGCGTGTTCAGACAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT----- 998
QY 379 GluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyVal 398
Db 999 GAGCTT-----GATGTCATGCGCACTGCGCTATCTATC 1031
QY 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1032 CAAAGCAGCGTCTCTGGAAACAATAACGGG-----CGGTACAACGCTAGTCAATGGCA 1085
QY 417 SerProValValAlaGlyAlaValThrLeuLeuValSer 429
Db 1086 TCTCCGACGCTTCGCGGAGCGGTCTGTTGATCTTTCT 1124
```

RESULT 9

US-10-033-325-1

; Sequence 1, Application US/10033325

; Publication No. US2003007322A1

; GENERAL INFORMATION:

```
; APPLICANT: Poulouse, Ayrookaran J.
; APPLICANT: Schellenberger, Volker
; APPLICANT: Kellis, Jr., James T.
; APPLICANT: Paech, Christian
; APPLICANT: Nadherny, Joanne
; APPLICANT: Naki, Donald P.
; APPLICANT: Collier, Katherine D.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Baeck, Andre C.
; TITLE OF INVENTION: Multiply-Substituted Protease Variants
; FILE REFERENCE: GC502-2-C1
; CURRENT APPLICATION NUMBER: US/10/033,325
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: US 09/178,155
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: US 08/956,323
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,564
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,324
; PRIOR FILING DATE: 1997-10-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Bacillus amyloliquefaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (96)...(1245)
; US-10-033-325-1

Alignment Scores:
Pred. No.: 2,12e-25 Length: 1497
Score: 331.00 Matches: 124
Percent Similarity: 40.40% Conservative: 59
Best Local Similarity: 27.37% Mismatches: 146
Query Match: 5.89% Indels: 124
DB: Gaps: 19

US-09-830-837-6 (1-1052) x US-10-033-325-1 (1-1497)
QY 8 LeuLeuLeuValValLeuLeuCysGlyLysHisLeuGlyAspArgLeuGluLys 27
Db 45 ATAATCTGTCTATGTTATCTGCAATCAATCAAAAGGAGAGGATAAGAGTGAGAGC 104
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 105 AAAAAAGTATGGATCATGTTGCTGTTTGTAGCG-----TTA 143
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 144 ATCTTTACGATGCGGTTCGCGACACATCTCTCCCGCGCGGGAATCAACGGG 203
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 204 GAAAGAAATATATGTCGGGTTT-----AAACAGACAATGAGCAGC 245
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsn 92
Db 246 ATGAGCGCGCTAAGAAGAAGATGTCATT----- 275
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 276 -----TCTGAAAAGGC 287
QY 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArgLys 131
Db 288 GCGAAGTGTAAAAGCAATTCAAATATGTAGACGAGCTTCAGTCACATTAAACGAAAA 347
QY 132 ValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArg 151
Db 348 GCTGTAAAGAAATGTAAA-----AAAGACCGGAGCGTCCT----- 383
```

```
QY 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySer 171
Db 383 -----
QY 172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuAlaIleProArg 191
Db 384 -----TACGTTGAAGAGATCAGTACAGTACGTCAGCGAGTCCGTCCTTAC 434
QY 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrThrGlyAlaAsn 211
Db 435 GCGGTATACAA--ATTAAGCCCTGCTCTGCTCAAGCTCTCAAGCTACATCGATCAAT 491
QY 212 ValArgValAlaValPheAspThrGlyLeuSerGlnLysHisProHisPheLys----- 229
Db 492 GTTAAGTAGCGGTATTCACAGCGGTATCGATCTCTCATCTGATTTAAAGGTAGCA 551
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
Db 552 ACGGAGCCAGCATGTTCTTCTGAACAAAT-----CCTTCCACAGAC 596
QY 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 597 ACAACTCTCACGGAACCTCACGTCGCGGCACAGTTCGGGCTCTTAATACTCAATCGGT 656
QY 263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
Db 657 GTATTAGCGGTTCGCCAAGCGCATCTTTAGCTGTAAAGTTCTCGGTGTCACGGT 716
QY 283 ValSerTrpThrPheLeuAspAlaPheAsnTrpAlaIleLeuLysLysIleAsp 302
Db 717 TCGGCCCAATACAGCTGGATCATTAAGGAATCGAGTGGCGATCGCAACAATATGGAC 776
QY 303 ValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAsp 320
Db 777 GTATTAACTAGAGCTCGCGGACCTCTGCTGCTGCTTTAAAGGGCGAGTGTAT 836
QY 321 LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly 340
Db 837 AAAGCC-----GTTGCTCCGCGCTGCTAGTGTGCGGCGCGCGTAAACGAGGC 887
QY 341 Pro-----LeuTrpGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
Db 888 ACTTCCGCGCAGCTCAGCACAGTGGCTACCTGCTGTAATACCTTCTGTCATTCAGTA 947
QY 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrp 378
Db 948 GCGCTGTGACAGCAGCAACCAAGAGCATCTTCTCAAGCGTAGACCT----- 998
QY 379 GluLeuProGlyGlyTrpGlyArgMetLysProAspIleValThrTrpGlyAlaGlyVal 398
Db 999 GAGCTT-----GATGTCATGGCAGCCTGCGGTATCTATC 1031
QY 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1032 CAAGCAGCGTCTCTGGAAACAAATACGG-----GCGTACACGCTAGTCAATGGCA 1085
QY 417 SerProValAlaGlyAlaValThrLeuLeuValSer 429
Db 1086 TCTCCGCGAGTTCGCGAGCGCTGCTTTCATCTTTCT 1124

RESULT 10
US-09-060-854B-1
; Sequence 1, Application US/09060854B
; Patent No. US20020081703A1
; GENERAL INFORMATION:
; APPLICANT: Estal, David Aaron
; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical
; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002008
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: GC532
; CURRENT APPLICATION NUMBER: US/09/060.854B
; CURRENT FILING DATE: 1998-04-15
```

```
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: B. amylioliquefaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (96)...(1245)
US-09-060-854B-1

Alignment Scores:      2,12e-25      Length:      1497
Pred. No.:            331,00         Matches:    124
Score:                40,40%         Conservative: 59
Percent Similarity:   27,37%         Mismatches:  146
Best Local Similarity: 5,89%         Indels:      124
Query Match:          10             Gaps:        19

US-09-830-837-6 (1-1052) x US-09-060-854B-1 (1-1497)

QY 8 LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys 27
Db 45 ATAATCTGCTATGTGTTATCTGCAATGAAAAAAGAGAGAGATAAGAGTGGAGGC 104
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 105 AAAAAGATGATGATCAAGTTTGTCTTGTCTTTAGCG-----TTA 143
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 144 ATCTTTAGCATGGCGTTCGGCAGCACATCTCTCCAGCGCGAGGAAATCAACAGGG 203
QY 53 GluTrpGluTrpIleValAlaPheAsnGlyTrpPheThrAlaLysAlaAsnSerPhe 72
Db 204 GAAAGAAATATATTCGCGGTT-----AAACAGCAATGAGCAGC 245
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn 92
Db 246 ATGACGCCCTAAGAAAGATGTCATT----- 275
QY 93 ProSerSerAspTrpProSerAspPheGluValIleGlnIleLysGluLysAla 112
Db 276 -----TCTGAAAAAGGC 287
QY 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArgLys 131
Db 288 GGGAAAGTGCAAAAGCAATCAATATGTAGCCAGCTTCAGTCACTTAACGAAAAA 347
QY 132 ValPheArgSerLeuLysTrpAlaGluSerAspProThrValProCysAsnGluThrArg 151
Db 348 GCTGTAAAGAAATTTGAAA-----AAAGACCCGAGCGTCGCT----- 383
QY 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySer 171
Db 383 ----- 383
QY 172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuAlaIleProArg 191
Db 384 -----TACGTTGAAGAGATCAGTACAGTACGTCAGCGAGTCCGTCCTTAC 434
QY 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrThrGlyAlaAsn 211
Db 435 GCGGTATACAA--ATTAAGCCCTGCTCTGCTCAAGCTCTCAAGCTACATCGATCAAT 491
QY 212 ValArgValAlaValPheAspThrGlyLeuSerGlnLysHisProHisPheLys----- 229
Db 492 GTTAAGTAGCGGTATTCACAGCGGTATCGATCTCTCATCTGATTTAAAGGTAGCA 551
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
Db 552 ACGGAGCCAGCATGTTCTTCTGAACAAAT-----CCTTCCACAGAC 596
```



```
Qy 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 597 AACAACTCTCAGGAACACTACGTTGCGGCACACGTCCTTAATCACTCAATCGGT 656
Qy 263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
Db 657 GTATTAGGCGTTCCGCAAGCGCATCATTACGCTGTAAAGCTTCCGGTGTGACGGT 716
Qy 283 ValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAsp 302
Db 717 TCCGGCCAAATACAGCTGGATCAATAACGAATCGATGGCGGATCCAAACATATGGAC 776
Qy 303 ValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAsp 320
Db 777 GTATTAAATAGATGACCTCGCGGACCTTCTGCTCTCTTAAAGCGGCGAGTTGAT 836
Qy 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
Db 837 AAAGCC-----GTTGATCCCGGCGTGTAGTCTGTTGCGGCGACCGCGTAACGAAGC 887
Qy 341 Pro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
Db 888 ACTTCGGCGGAGCTCAAGCAGTGGGCTACCTCTGGTAAATACCTTCTGTCAATGCGAT 947
Qy 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrTrp 378
Db 948 GCGCTGTGTACAGCAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT----- 998
Qy 379 GluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyVal 398
Db 999 GAGCTT-----GATGTCATGCGACCTGGCGTATCATC 1031
Qy 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1032 CAAGCAGCGTCTCTGGAAACATACGG-----GCGTACACGCTAGCTCAATGGCA 1085
Qy 417 SerProValValAlaGlyAlaValThrLeuLeuValSer 429
Db 1086 TCTCCGACGCTTCCGCGAGCGGCTGCTTTGATTCTTTCT 1124
```

RESULT 11

```
US-10-090-624-11
; Sequence 11, Application US/10090624
; Patent No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090, 624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11
```

```
Alignment Scores:
Pred. No.: 4,99e-24 Length: 1977
Score: 320.00 Matches: 127
Percent Similarity: 40.72% Conservative: 64
Best Local Similarity: 27.08% Mismatches: 185
```

```
Query Match: 5,70% Indels: 93
DB: 12 Gaps: 22
US-09-830-837-6 (1-1052) x US-10-090-624-11 (1-1977)
Qy 199 AlaAspValLeuTrp---GlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPhe 217
Db 439 GCGATACCTCTGGAACCTCCCTCGGTACGAGGAAGGTGTGGTGTGCATCGTC 498
Qy 218 AspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrp 237
Db 499 GATACGGGTATAGACGGGAACACCCCGATCTGAAGGC-----AAGGTCATAGGCTGG 552
Qy 238 -----ThrAsnGluArgThrLeuAspGlyLeuHisGlyThrPheVal 253
Db 553 TAGCAGCGCTCAACGGCAGTCGACCCCTACGATGACCGAGGACACGACCGACGTT 612
Qy 254 AlaGlyValIleAlaSerMetArgGluCysGln-----GlyPheAlaProAsp 269
Db 613 GCGGGTATCTTCCGGAACCGGACGCTTAACCTCCAGATACATAGCGCTCGCCCGGC 672
Qy 270 AlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPhe 289
Db 673 GCGAAGCTCTCGGGTCAAGGTTCTCGGTGCGCAGCGTTCCGGGAAGCGTCTCCACCATC 732
Qy 290 LeuAspAlaPheAsnTyrAlaIleLeuLysLys-----IleAspValLeuAsn 305
Db 733 ATCGCGGTCTTGACTGGGTCTGTCAGCAACAGGACAAAGTACGGGATAAGGTCATCAAC 792
Qy 306 LeuSerIleGly-----GlyProAspPheMetAspHisProPheVal 319
Db 793 CTCTCCCTCGCTCTCCAGAGCTCGAGGACCGACGACTCCCTCAGTCAGCGC---GTC 849
Qy 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
Db 850 AACACGCTCGGAGCGC-----GGTATAGTACTGCTGCGTCCGCGCGGCAACAGC 900
Qy 340 GlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
Db 901 GGGCGGAACACCTACACCGCTCGGCTCACCCGCGCGAGCAAGGTTCATAACCGTCGT 960
Qy 360 GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu 379
Db 961 GCAGTTGACAGCAACGACACATCGCAGCTTCTCCAGCAGGGA----- 1005
Qy 380 LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyVal--- 398
Db 1006 ---CCGACCGCGGAGGAGGCTCAAGCGGGAAGTCTGTCGCCCGCGGTGACATCATA 1062
Qy 399 -----ArgGlySerGlyValLysGlyGlyCys-----ArgAlaLeu 410
Db 1063 GCGCGCGCGCGAGCGGAACACGATGGGCAACCCCGGATAACGACTACTACCAAGGCC 1122
Qy 411 SerGlyThrSerValAlaSerProValAlaGlyAlaValThrLeuLeuValSerThr 430
Db 1123 TCTGGAACACGATGCGCCACCGCGAGCTTTCGGCGGTTCGCGCGCTCATC----- 1173
Qy 431 ValGlnLysArgGluLeuValAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAla 450
Db 1174 CTCAGCGCCACCGGAGCTGGACCCCGGAAGGTGAAGACCGCCCTCATCGAGACGCC 1233
Qy 451 ArgArgLeuProGlyValAsnMetPheGlu-----GlnGlyHisGlyLysLeuAspLeu 468
Db 1234 GACATAGTCGCCCCCAGGAGATAGCGGACATCGCTACGTCGCGGTAGGGTAGAGTC 1293
Qy 469 LeuArgAlaTyrGlnIleLeuAsnSerTyrLysProGlnAlaSerLeu-----SerPro 486
Db 1294 TACAAGGCCATCAAG-----TACGAGCACTACGCCAAGCTCACTTCACCGGC 1341
Qy 487 SerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSerGlnProIleTyr 506
Db 1342 TCGTCCGCGGACAAAGGGAAGCGCCACCCACACCTTCGACGTCAGCGC----- 1389
```

507 TyrGlyMetProThrValValAsnValThrIleLeuAsnGlyMetGlyValThrGly 526
1390 -----GCCACCTTCGTGACCGCCACCTCTACTGGGACACGGGC----- 1428
527 ArgIleValAspLysProAspTyrGlnProTyrLeu---ProGlnAsnGlyAspAsnIle 545
1429 -----TCGAGCGACATCGACCTCTACCTCTACGACCCCAACGGAACGAGGT 1476
546 GluValAlaPheSer-----TyrSerSerValLeu 555
1477 GACTACTCTACACCGCTTACGAGGCTTCGAGAGGCTGCTACTACAACCGACCGCC 1536
556 TrpProTyrSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaLaserTyr 575
1537 GGAACTGG-----ACGGTCAAGGTCGTACGCTAC 1566
576 GluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThrGluSer 595
1567 AGGCGCGCGG-----AACCACGAGTTCGAGTTCGCGGAGGAGGCTCAGCCAG 1620
596 LysAsnGly-AlaGluInThrSerThrValLysLeuProIleLysValLysIleIlePr 615
1621 TCGGCGCGCGCAACCGAATCCAAACC-----CCAAACCGAACCACCGCGGCC 1671
615 oThrProArgSerLysArgVal 623
1672 ACCGACACCGACCTTCACCGGTT 1696
RESULT 12
US-10-090-624-2
; Sequence 2, Application US/10090624
; Patent No. US2002013235A1
GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Kiyoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; CURRENT FILING DATE: 2002-03-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2
Alignment Scores:
Pred. No.: 2, 02e-22 Length: 1236
Score: 302.00 Matches: 119
Percent Similarity: 43.78% Conservative: 64
Best Local Similarity: 28.47% Mismatches: 150
Query Match: 5.38% Indels: 87
DB: 12 Gaps: 19
US-09-830-837-6 (1-1052) x US-10-090-624-2 (1-1236)
QY 186 LeuArgAlaIleProArgGlnValAlaGlnThrIleLeuGlnAlaAspValLeuTrpGlnMet 205
DB 7 TTAAAGGACTGGATGAGTCTGACGCTCAAGTTATG---GCAACTAGCTTTGGAATTG 63
QY 206 GlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHis 225
DB 64 GGATATGATGGTCTTGGAATCAATAGGAATAATTCACACCTGGAATTCAGCTCTCTCAH 123

QY 226 ProHisPhelys---AsnValLysGluArgThrAsnTyrThrAsnGluArgThrLeuAsp 244
DB 124 CCAGATTCACAGAAAGATAATGGTGGGTAGATTTCTCAATGGTAGGAGTTATCCA 183
QY 245 ---AspGlyLeuGlyHisGlyThrPheValAlaGlyVal-----IleAla 258
DB 184 TAGCATACCATCATGATGATGCTTCAATAGCAGCTGGTACTTGGAGCAGCA 243
QY 259 SerMetArgLysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPhe 278
DB 244 AGTAATGGCAAGTACAGGAATGGCTCCAGGAGCTAAGCTGGCGGGAATTAAGGTTCTA 303
QY 279 ThrAsnAsnGlnValSerTyrThrSerThrPheLeuAspAlaPheAsnTyrAlaIleLeu 298
DB 304 GGTGCGGATGTTCTGGAAGCATATCTACTATAATTAAGGAGTGTGAGTGGCGGTTGAT 363
QY 299 LysLys-----IleAspValLeuAsnLeuSerIleGly----- 309
DB 364 AACAAGATAAGTACGGAATTAAGTGCATTAAATCTTCTCTGTTCAAGCCAGAGCTCA 423
QY 310 ---GlyProAspPheMetAspHisProPheValAspLysValTyrGluLeuThrAlaAsn 328
DB 424 GATGCTACTGACGCTCTAAGTCAGCT---GTTAATGACGCTGGGATGCT--- 471
QY 329 AsnValIleMetValSerAlaIleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsn 348
DB 472 GGATAGTGTGTGTGGTGGCTGGAAACAGTGCAGCTAACAGATATACATCGGTTCT 531
QY 349 ProAlaAspGlnMetAspValIleGlyValGlyIleAspPheGluAspAsnIleAla 368
DB 532 CCAGCAGCTGCAAGCAAGTTATACAGTGGCCCTGCACAGTAGTGTGTATACAA 591
QY 369 ArgPheSerSerArgLysMetThrThrTrpGluLeuProGlyGlyTyrGlyArgMetLys 388
DB 592 AGCTTCTCAAGCAGAGGG-----CCAACTCAGACGGCAGGCTTAAG 633
QY 389 ProAspIleValThrTyrGly-----AlaGlyValArgLysSerGlyValLysGly 405
DB 634 CCTGAGGTGTGTGCTCCAGGAACCTGGATAATTCTCCAGACAGTGGAGTAAGCATG 693
QY 406 Gly-----CysArgAlaLeuSerGlyThrSerValAlaSerProVal 419
DB 694 GGTCACCAACCAATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGCACTCCTAC 753
QY 420 ValAlaGlyValAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnPro 439
DB 754 GTAGCTGTATTGTCAGCCCTCTTG-----CTCCAAGCAGCACCCGAGCTGGAGCTCA 804
QY 440 AlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMetPhe 459
DB 805 GACAAGTAAACACAGCCCTCATAGAAACTGCT----- 837
QY 460 GluGlnGlyHisGlyLysLeuAspLeuArgAlaTyrGlnIle----- 474
DB 838 -----GATATCGTAAGCCAGATGAATACCGCATATACCTTAC 876
QY 475 -----LeuAsnSerTyrLysProGlnAlaSerLeuSerProSerTyrIleAsp 490
DB 877 GTGCAGGTAGGTTAATGCATACAAG-----GCTATTAACCTACCATAC 921
QY 491 LeuThrGluCysProTyrMetThrProTyrCysSerGln-----ProIle 505
DB 922 TATGCAAAAGCTAGTGTTCAC-TGGATATATTGCCAACAAGGCGACCAACTCACCAGTT 980
QY 506 TyrTyrGlyGlyMetProThrValValAsnValThrIleLeuAsnGlyMetGlyValThr 525
DB 981 CGTTAT-----TAGCGGAGCTTCTGCTGTAACCTGCCACATATTACTG 1022
QY 526 GlyArgIleValAspLysPro-AspTyrGlnProTyrLeuProGlnAsnGlyAspAsnIle 545
DB 1023 GGACATGCCATAGCAGCTTCTATCTTACCTCTACGATCC---AATGGAAACAGGT 1079

Qy 545 eGluValAlaPheSerTyrSerValLeuTyrProTyrSerGlyTyrLeu 562
Db 1080 TGAC-----TACTCTTACACCGC-CTACTATGATTCGAAAGGTTGGTTA 1124

RESULT 13

US-10-090-624-15

; Sequence 15, Application US/10090624

; Patent No. US20020132335A1

; GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru

; APPLICANT: MORISHITA, Mio

; APPLICANT: SHIMOJO, Tomoko

; APPLICANT: ASADA, Kiyozo

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

; FILE REFERENCE: TAKAKURA-6

; CURRENT APPLICATION NUMBER: US/10/090,624

; PRIOR FILING DATE: 2002-03-06

; PRIOR APPLICATION NUMBER: 09/445,472

; PRIOR FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 15

; LENGTH: 1962

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Synthetic

; US-10-090-624-15

Alignment Scores:

Pred. No.:	4.21e-22	Length:	1962
Score:	302.00	Matches:	119
Percent Similarity:	43.78%	Conservative:	64
Best Local Similarity:	28.47%	Mismatches:	150
Query Match:	5.38%	Indels:	87
DB:	12	Gaps:	19

US-09-830-837-6 (1-1052) x US-10-090-624-15 (1-1962)

Qy 186 LeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTyrGlnMet 205
Db 403 TTAAAGGACTGGATGAGTCTGCAGCTCAAGTTATG---GCAACTTACGTTTGGAACTTG 459
Qy 206 GlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHis 225
Db 460 GGATATGATGGTTCTTGGAAATCAACAATAGGAATATTGACACTGGAAATTCAGCTTCTCAT 519
Qy 226 ProHisPheLys---AsnValLysGluArgThrAsnTyrThrAsnGluArgThrLeuAsp 244
Db 520 CCAGATCTCAAGGAAAGTAATGGTGGGTAGATTTTGTCAATGGTAGAGTTATCCA 579
Qy 245 ---AspGlyLeuGlyHisGlyThrPheValAlaGlyVal-----IleAla 258
Db 580 TACGATGACCATGACATGCACTGATGATGCTCAATAGCAGCTGCTACTGGAGCAGCA 639
Qy 259 SerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPhe 278
Db 640 AGTAATGGCAAGTACAAGGAAATGGCTCCAGAGAGCTAAGTCGGCGGAATTAAGGTTCTA 699
Qy 279 ThrAsnAsnGlnValSerTyrThrSerTyrPheLeuAspAlaPheAsnTyrAlaIleLeu 298
Db 700 GGTCCGATGGTCTGGAACATATCTACTATTAATTAAGGAGTTGAGTGGCGGCTGAT 759
Qy 299 LysLys-----IleAspValLeuAsnLeuSerIleGly----- 309
Db 760 AACAAAGATAAGTACGGAATTAAGTCTAATTAATTTCTTCTGTGTTCAAGCCAGAGCTCA 819
Qy 310 ---GlyProAspPheMetAspHisProPheValAspLysValTyrGluLeuThrAlaAsn 328
Db 820 GATGGTACTACGCTCTAAGTCAGGCT---GTTAATGCAGCTGGGATGCT----- 867

SEQ ID NO 5
LENGTH: 4765
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-090-624-5

Alignment Scores:

Read No.: 1,14e-19 Length: 4765
Score: 285.00 Matches: 185
Percent Similarity: 33.44% Conservative: 116
Best Local Similarity: 20.56% Mismatches: 313
Query Match: 5.07% Indels: 286
Gaps: 39

US-09-830-837-6 (1-1052) x US-10-090-624-5 (1-4765)

```
.OY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAla----- 258
|||||
Db 1204 GATGGTCAGGTCACGGAACCTCACGTAGCTGGAACCTGCTGGTTAGCAGACAACAAAT 1263
OY 258 ----- 258
Db 1264 GATGCTGGGATGGCTCAGTATGCTCTGGTGAATGGGAAGTGTCTCAAGACTCTAT 1323
OY 259 -----SerMetArgGluCysGlnGlyPheAlaProAspAlaGlu 271
|||||
Db 1324 GATGGGATATACGAACGTTTACACAGACACCGGTCAGGGTGTGCTCCAGGTGCCAA 1383
OY 272 LeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAsp 291
|||||
Db 1384 ATAATGGCAATAAGAGTTCTTAGGAGTGTAGGACGGGTAGCATGCTGGGATATATAGAA 1443
OY 292 AlaPheAsnTyrAlaIleLeuLysLysIleAspValLeuAsnLeuSerIleGlyGlyPro 311
|||||
Db 1444 GGTATGACATACGACGACCAACCCATGTCGACAGCTTATAAGCATGAGTCTCGGTGA--- 1500
OY 312 AspPheMetAspHisProPheValAsp-----LysValTrpGlu 324
|||||
Db 1501 -----AATGCTCCATACATAGATGGTACTGATCCAGAAAGGTTGCTGGTATGAG 1551
OY 325 LeuThrAlaAsn---AsnValIleMetValSerAlaIleGlyAsnAspGlyProLeuTyr 343
|||||
Db 1552 CTTACCGAAGTACGCTGTGTATCTGTAATAGCTGCAGGAATGAAGTCTCGCAT 1611
OY 344 GlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly----- 359
Db 1612 AACATGTTGGAGTCTGTGTGTGCAACAAGGCATACCTGTGGAGCTGTCAGTG 1671
OY 360 -----GlyIleAspPheGluAsp----- 365
Db 1672 CCCATTACGTTGGAGTTTATGTTTCCCAAGCACTGGATATCCTGATCTATGGATTC 1731
OY 366 -----AsnIleAlaArgPheSerSerArgGlyMetThr 376
|||||
Db 1732 TATTACTTCCCGCTACACAACGTTTGAATAGCATCTTCTCAAGCAGAGG----- 1785
OY 377 ThrTrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAla 396
|||||
Db 1786 -----CCGAGATAGATGTGTAATAAAACCAATAGTAGTGGCTCCAGGTTAC 1833
OY 397 GlyValArgGlySerGlyValLys-----GlyGlyCysArgAlaLeuSerGlyThrSer 414
|||||
Db 1834 GGAATTTACTCATCCCTCGCGATGGTGGATGGCGGAGCTGACTTCTATGCTGGAACCTCG 1893
OY 415 ValAlaSerProValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArg 434
|||||
Db 1894 ATGCTATCCACATGTCACGGTGTGCTGCACTCTCATAGCGGGGCAAGCGCGAG 1953
OY 435 GluLeu---ValAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeu 453
|||||
Db 1954 GGAATATACTACAATCCAGATATATTAAGAGGTTCTTGAGAGCGGTGCACCTGGCTT 2013
```

```
OY 454 ProGly-----ValAsnMetPheGluGlnGlyHisGlyLysLeu 466
|||||
Db 2014 GAGGAGATCCATATATACTGGCAGAAAGTACACTGAGCTTGACCAAGGTCATGGTCTTGT 2073
OY 467 AspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrIleProGlnAlaSerLeuSerPro 486
|||||
Db 2074 AACGTTACCAAGTCTGGGAAATCCTTAAGCT----- 2106
OY 487 SerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSerGlnProIleTyr 506
|||||
Db 2107 -----ATAAACGCGCACCACTCTCCCAATTGTTGATCCTGGCGAGACAAGTCC----- 2154
OY 507 TyrGlyGlyMetProThrValValAsnValThrIleLeuAsnGlyMetGlyValThrGly 526
|||||
Db 2155 TACAGCGACTTTCGGGAGTACTTGGTGTGACGCTTATAAGAGGTCTCTAGCAAGGAAC 2214
OY 527 ArgIleValAspLysProAspTrpGlnProTyrLeuProGlnAsnGlyAspAsnIleGlu 546
|||||
Db 2215 TCTATACCTGACATGTCGAGTGG-----CACATTAAGTACGTAGGGGACACGAGTAC 2268
OY 547 ValAlaPheSer---TyrSerValLeuTrp-----ProTrp---SerGlyTyrLeu 562
|||||
Db 2269 AGAATCTTTGAGATCTATGCACTGAGCCATGGATTAAGCTTTTGTTCAGTGAAGTGA 2328
OY 563 AlaIleSerIleSer-----ValThrLysLysAlaAlaSerTrpGluGlyIleAla 579
|||||
Db 2329 ATTCAGAGAACAAATACCGAGTGTGCTTAGGTGAATATGATGATAGGCTCTGAG 2388
OY 580 GlnGlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAla 599
|||||
Db 2389 CCAGGCTCTCTATGTTGGAAGGATATCATGATGATCCCAACACGCCAGTATTTAGAGAC 2448
OY 600 GluGlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProArg 619
|||||
Db 2449 GAGATCTTGAACAATTTGTTATCCGAGAAGTTC-----ACTCCTGAGAAC 2496
OY 620 SerLysArgValLeuTrpAspGln-----TyrHisAsnLeu 631
|||||
Db 2497 AATTACACCTCACCCTGGTATGATATTAATGTCGAGAAATGGTGTGACTCACCACTCTTC 2556
OY 632 ArgTyrProGly----- 636
Db 2557 ACTGTGCTGAGGAGTGGAGCTTCTTACGCGATGACACATCTACCTGGGACTACGGTCTG 2616
OY 637 -----TyrPheProArg----- 640
Db 2617 TACAGACCATGGAATGTTTGTGTTCCCATACACGCTAGATTATCTCCGCTGCAGTC 2676
OY 641 AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn 660
Db 2677 TCATTCATCCCTGCGAACTGGGAGCTAGTATGGACTGGA-----TTTAAC 2724
OY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
|||||
Db 2725 TTTGCACCCCTCTATGAG-----TCGGGCTTCTTGTAGGATTTACGGAGTA 2772
OY 681 ProPheThrCys----- 684
Db 2773 GAGATAACTCCCAAGCGTTTGGTATACATTAACAGGACATACCTTGACACTAACCTGAATC 2832
OY 685 -----PheAspAlaSerGlnTyr-----GlyThrLeuLeuMetVal 696
|||||
Db 2833 TCAATTGAATTCATATTAATCTATGCCCCCAATTAATGCAACTCTAATCCCACT 2892
OY 697 Asp-----SerGluGluGluTyrPhe----- 703
Db 2893 GGCCTTGGAACTACAAATGCGAGCGTTGAAAGCGTTGGTGTGATGAGAGATTTCTTATAAG 2952
OY 704 -----ProGluGluIleAlaLysLeuArgAspValAspAsn----- 716
|||||
Db 2953 GGCATTGAGTTCCTGAAGGCCCGCAGAGTTGAAGATTAGGATAGCAACCAAGTGT 3012
```

```
QY 717 -----GlyLeuSerLeuValIlePheSerAspTrpTyrAsn----- 728
Db 3013 CCGAATTCAGATCAGACTGTACCTTTATGACAGTAAGGCAATTTAGTGGCTTAGAT 3072
QY 729 -----Thr 729
Db 3073 GGAACCCACACAGCAGAAGAGGTGTAGTATCCTTAAGCTGGAGTTTATCA 3132
QY 730 SerValMetArgLysValLysPheTyrAspGluAsnThrArg----- 743
Db 3133 ATAGTAGTACATCGTTACAGCGTCAGGACGAAATGCTTAATCCACGACCAACCCTTT 3192
QY 744 -----GlnTrpTrpMetProAspThrGlyAlaAsn----- 754
Db 3193 GACTTAGTGTCAATGACCCCT---GATAATGGAACATAAAGCTTGACAAAGACTCG 3249
QY 755 ---IleProAlaLeuAsnGluLeuLeuSerValTrp-----AsnMetGly 768
Db 3250 ATTATTCTTGGAGCAATGAAGCGTAGTGTGAACCTGCAACATAACAATTGATAGAGAT 3309
QY 769 PheSerAspGlyLeuTyrGluGlyPheThrLeuAlaAsnHisAspMetTyrTyrAla 788
Db 3310 CATCTACAGGAGTATACCTGTGTATCATAGATTAAGATTAATGAGGTCTACAGGAT 3369
QY 789 SerGlyCysSerIleAlaLysPheProGluAspGlyValIleThrGlnThrPhe--- 807
Db 3370 ACAATACTTCAATTGCGAAATACCCATAACTTGGTAATGACAAGCGGACTTTGCC 3429
QY 808 -----LysAspGlnGlyLeuGluValLeuLys 816
Db 3430 GTTGGTCTCACACCACAGCAGGAGTACTTGGAGAGCTAGAAATACACTCTAATGTGA 3489
QY 817 GlnGluThrAlaValGluAsnValPro-----IleLeuGlyLeuTyrGln 832
Db 3490 AGCATGCCCCATACACTAGAGCTGTGCCAAATGCTACAGTATATAGGAACACTACACC 3549
QY 833 IleProAlaGluGlyGlyArgIleValLeu---TyrGlyAspSerAsnCysLeuAsp 851
Db 3550 TACCTCACAGACGAAAAACGGTACAGTGCACATTCACGTATGCTCCAACTAAGTTAGGCAGT 3609
QY 852 AspSerHisArg-----GlnLysAspCysPheTrpLeuLeuAspAlaLeuGln 868
Db 3610 GATGAATCAGCTCATAGTATGTAAGAAGAGAACTTCAACACATTAAGAGAACCTTCCAA 3669
QY 869 -----TyrThrSerTyrGlyValThrProProSerLeuSerHis 881
Db 3670 ATCAGATATCAGAGCTGNAATACTGAAGAGCACATAAATCAGCCCAAGCTTGCAATG 3729
QY 882 SerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMetGlu 901
Db 3730 TCATCA-----CCAGAAGCAAAATGCTACCATAGTATCATCAGTTGAGATGGAG 3774
```

RESULT 15

```
US-08-322-678-11
; Sequence 11, Application US/08322678
; Publication No. US20030077807A1
; GENERAL INFORMATION:
; APPLICANT: Gravyar, Thomas P
; APPLICANT: Bott, Richard R
; APPLICANT: Wilson, Lori J
; TITLE OF INVENTION: Subtilisin Variants
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc
; STREET: 180 Kimball Way
; CITY: So. San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,678
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC235-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-322-678-11

Alignment Scores:
Pred. No.: 2,49e-20 Length: 1140
Score: 282.00 Matches: 113
Percent Similarity: 40.50% Conservative: 64
Best Local Similarity: 25.86% Mismatches: 166
Query Match: 5.02% Indels: 94
Db: 7 Gaps: 17
```

```
US-09-830-837-6 (1-1052) x US-08-322-678-11 (1-1140)

QY 41 LeuThrLeuLysValGluPheSerSerThrVal-----ValGluTyr 54
Db 40 CTACTCATTTCTGCTTCTTTTACTTCATCGATCGCTGCTGCTGAAGAAGCAAGAA 99
QY 55 GluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSer 74
Db 100 AATATTATTATTTGGCTTTTAAAT-----GAGCAGGAAGCTGTCAGTGAGTTGTAGAA 150
QY 75 SerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsnProSer 94
Db 151 CAAGTAGAGGCAATACGAGGTC-----GCCATCTCTCTGAGGA----- 192
QY 95 SerAspTyrProSerAspPheGluValIleGlnIleLysGlyLysGlnLysAlaGlyLeu 114
Db 193 -----CAGGAAGTCGAAATTTGAATTCCTTCATCAATTTGAACGATTTCCGTGT 240
QY 115 LeuThrLeuGluAspHisPro-----AsnIleLysArgValThrProGlnArgLys 131
Db 241 TTATCGTTTGAGTTAAGCCCAAGAGATGTGACGCGCTTGAACCTCGATCCA----- 291
QY 132 ValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArg 151
Db 292 -----GCGATTTCTTATATTGAAGAGGATGCAGAGTA-----ACGACA 330
QY 152 TrpSerGlnLys-----TrpGlnSerSerArgProLeuArgAlaSerLeuSerLeu 169
Db 331 ATGGCGCAATCAGTCGCAATGGGCAATTAGCCGT----- 363
QY 170 GlySerGlyPheThrPheHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIle 189
Db 363 ----- 363
QY 190 ProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGly 209
Db 364 -----GTGCAAGCCCCAGCTGCCCATACCGTGGATTTGACAGGT 402
QY 210 AlaAsnValArgValAlaValPheAspThrGlyLeuSerGlyLysHisProHisPheLys 229
Db 403 TCTGGTGTAAAGCTTGCTGCTCCGATACAGGATTTTCC---ACTCATCCAGACTTA--- 456
QY 230 AsnValLysGluArgThrAsnTrpThr---AsnGluArgThrLeuAspAspGlyLeuGly 248
Db 230 AsnValLysGluArgThrAsnTrpThr---AsnGluArgThrLeuAspAspGlyLeuGly 248
```

```

; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC235-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-322-678-12
Alignment Scores:
Pred No.: 5,23e-20 Length: 1140
Score: 279.00 Matches: 113
Percent Similarity: 40.50% Conservative: 64
Best Local Similarity: 25.86% Mismatches: 166
Query Match: 4.97% Indels: 94
DB: Gaps: 17
US-09-830-837-6 (1-1052) x US-08-322-678-12 (1-1140)
QY 41 LeuThrLeuLysValGluPheSerSerThrVal-----ValGluTyr 54
DB 40 CTACTCAATTCCTGCTTGTATTTAGTTCATCGATCGCATCGCTCGCTGAAGAAGCAAAAGAA 99
QY 55 GluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSer 74
DB 100 AAATATTAAATGGCTTTAAT-----GAGCAGGAAGCTGTCAGTGGATTTGTAGAA 150
QY 75 SerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnProSer 94
DB 151 CAAGTAGAGGCAATACGAGGTC-----GCCATTCTCTCTCAGGAA----- 192
QY 95 SerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeu 114
DB 193 -----GAGGAAGTCGAATTTGAATTTGCTTCATGAATTTGAAACGATTCCTGTT 240
QY 115 LeuThrLeuGluAspHisPro-----AsnIleLysArgValThrProGlnArgLys 131
DB 241 TTATCGTTGAGTTAAGCCCAAGATGTGGACGCGCTTGAACCTCATCA----- 291
QY 132 ValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArg 151
DB 292 -----GCGATTCTTATATTGACAGGATGCAGAATA-----ACGACA 330
QY 152 TrpSerGlnLys-----TrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeu 169
DB 331 ATGGCGCAATCAGTCGCATGGGAATTAGCCGT----- 363
QY 170 GlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuArgAlaIle 189
DB 363 ----- 363
QY 190 ProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGly 209
DB 364 -----GTGCAAGCCCACTGCCATCAACCGTGGATTCAGAGT 402
QY 210 AlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys 229
DB 403 TCTGCTGTAAAGTTCTGCTCTCGATACAGTATTTC-----ACTCATCCAGACTTA--- 456
QY 230 AsnValLysGluArgThrAsnTrpThr-----AsnGluArgThrLeuAspAspGlyLeuGly 248
DB 457 AATATTCGTGGTGGCGCTAGCTTTGTACAGGGAACCATTCCTCACTCAAGTGGGAATGGG 516
QY 249 HisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----CysGlnGly 265
DB 517 CATGCCACCATGTGGCCGGGAGCATTCCTGCTTTAGACAACTCGATGGCGTCTTGGC 576

```

RESULT 16

```

; Sequence 12, Application US/08322678
; Publication No. US20030077807A1
; GENERAL INFORMATION:
; APPLICANT: Graycar, Thomas P
; APPLICANT: Bott, Richard R
; APPLICANT: Wilson, Lori J
; TITLE OF INVENTION: Subtilisin Variants
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc
; STREET: 180 Kimball Way
; CITY: So. San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,678
; FILING DATE: 15-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```
QY 266 PheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyr 285
Db 577 GTAGCGCGGCGGCGGAACTATCGCTTTAAAGTATTAGGGCGGAGCGGCTTCAGGCGCC 636
QY 286 ThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLeuLysIleAspValLeuAsn 305
Db 637 ATCAGCTCGATTCCCAAGATTGGAATGGGAGGGAACAATGGCATGCACGTTGTGTAAT 696
QY 306 LeuSerIleGlyPheProAspPheMetAspHisProPheValAspLysValTrpGluLeu 325
Db 697 TTGAGTTAGGAAGCCCTTCG---CCAAAGTCCACACTTCAGCAAGCTGTATAGCGCG 753
QY 326 ThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyProLeuTyrGlyThr 345
Db 754 ACTCTAGAGCGCTTCTGTGTAGCGCATCTGGGAATTCAGGTGCA-----GGCTCA 807
QY 346 LeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyGlyIleAspPheGluAsp 365
Db 808 ATCAGCTATCCGCGCCGTTATGCGAAGCGAATGGCAGTCGGAGCTACTGACCAAAACAAC 867
QY 366 AsnIleAlaArgPheSerArgGlyMetThrTrpGluLeuProGlyGlyTyrGly 385
Db 868 AACCGCGCGAGCTTTTCACAGTATGCGCGAGGCGTT----- 903
QY 386 ArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGlySerGlyValLysGly 405
Db 904 -----GACATTTCGCACCGACGAGTGTAAAGTCGAGAGCACATACCCAGGTTCA 951
QY 406 GlyCysArgAlaLeuSerGlyThrSerValAlaSerProValValAlaGlyAlaValThr 425
Db 952 ACATATGCCAGCTTAACGGTATCATGATGGCTACTCCCTCATGTTGAGTGCAGCGGCC 1011
QY 426 LeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSerMetLysGlnAla 445
Db 1012 CTGTGT-----AAACAAAGAAGCCCATCTTGTGTCCTCAATGTACAAATCCGCAATCAT 1062
QY 446 LeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGluGlnGly 462
Db 1063 CTAAGAATACGGCAACGAGCTTAGGAAGCACCAACTGTATGGAAGCGGA 1113
```

RESULT 17

```
US-09-920-118-21
; Sequence 21, Application US/09920118
; Patent No. US20020102734A1
; GENERAL INFORMATION:
; APPLICANT: Menzel, Rolf
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIRECTED GENE ASSEMBLY
; FILE REFERENCE: 10424-003
; CURRENT APPLICATION NUMBER: US/09/920,118
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/222,134
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(576)
US-09-920-118-21
```

```
Alignment Scores:
Pred. No.: 1,34e-18 Length: 576
Score: 261.50 Matches: 63
Percent Similarity: 51.87% Conservative: 34
Best Local Similarity: 33.69% Mismatches: 77
Query Match: 4.66% Indels: 13
DB: 10 Gaps: 6
```

US-09-830-837-6 (1-1052) x US-09-920-118-21 (1-576)

```
QY 197 LeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaVal 216
Db 1 ATTAAGCGCGCGGCTCTTCACTCTCAAGGCTACACAGGTTCTTAACGTAAGTAAGCGCTA 60
QY 217 PheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsn 236
Db 61 ATTGACAGCGGAATGACTTCTCATCTCTGACTTG---AACGTCAGAGCGGAGCAAGC 117
QY 237 TrpThrAsnGluArgThr-----LeuAspAspGlyLeuGlyHisGlyThrPheValAla 254
Db 118 TTGCTACCTTCTGAACAAACCCATCAACAGATGGCAGTTCTCAGCGCACACATGTAGCC 177
QY 255 GlyValIleAlaSerMetArgGlu-----CysGlnGlyPheAlaProAspAlaGlu 271
Db 178 GGTAGGTTGCCGCACTTAATACTCAATCGGTGTTTTGGCGGTAGGCCCAACGCAATCG 237
QY 272 LeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAsp 291
Db 238 TTATATGACGATAAAGTTCTTGATTCAACAGGAACAGCGCAATACAGCTGGATTATTATAC 297
QY 292 AlaPheAsnTyrAlaIleLeuLysIleAspValLeuAsnLeuSerIleGlyGlyPro 311
Db 298 GGCATTGAGTGGGCCATTTCCAAACAAATGACGATTAACATGAGCCTTGGCGGACCT 357
QY 312 AspPheMetAsp-----HisProPheValAspLysValTrpGluLeuThrAlaAsn 329
Db 358 TCTGGTTTACAGCTTTGAAATCAGTCGTTGATAGAGCC-----GTAGCCAGCGGT 408
QY 330 ValIleMetValSerAlaIleGlyAsnAspGlyPro-----LeuTyrGlyThrLeuAsn 347
Db 409 ATCGTCGTTGTTGTCGAGCGCGGAATGAAGGCACTTCCGGAAGCTCAACACACAAATCGGC 468
QY 348 AsnProAlaAspGlnMetAspValIleGlyValGlyIleAspPheGluAspAsnIle 367
Db 469 TATCTCTCAAAATATCTTCTTACCATTCGCGTAGGTGCGGTAAACACACGCAACCAAGA 528
QY 368 AlaArgPheSerSerArgGly 374
Db 529 GGTTCATTCTCAAGCGGTAGT 549
```

RESULT 18

```
US-09-927-827-33
; Sequence 33, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseyer, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 33
; LENGTH: 3788
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2788)
US-09-927-827-33
```

```
Alignment Scores:
Pred. No.: 3,38e-17 Length: 3788
Score: 260.50 Matches: 111
Percent Similarity: 43.81% Conservative: 59
Best Local Similarity: 28.61% Mismatches: 124
Query Match: 4.64% Indels: 96
DB: 9 Gaps: 17
```

US-09-830-837-6 (1-1052) x US-09-927-827-33 (1-3788)


```
QY 223 uLysHisProHisPheLys-----AsnVa 231
      |||||
Db 641 GGATCATCCGGCTTGAACAAAGTACAGAGGCTACACCCCTTAAACCGAAGCAGTCC 700
      |||||
QY 231 lLysGluArgThrAsnTrp-----ThrAsnGluArgThrLeuAspAspGlyLe 247
      |||||
Db 701 AGACAATAGTTCAACTGGTTTGCATCGGTAAAGCAATAAGAAACCCATATGATGAGTT 760
      |||||
QY 247 uGlyHisGlyThrPheValalaGlyValIle-----AlaSerMetArgGluCy 263
      |||||
Db 761 GGCCACGACGACATGTCAGAGCAGCATGCTCGCTCAGAGCGCGGCGCAAAACCA 820
      |||||
QY 263 sGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThr----- 279
      |||||
Db 821 GATCGGTGTCGCCCGCAGCGCAATAGTATGTCAAGGCCTTTTCAGAGGACGCGG 880
      |||||
QY 280 -----AsnAsnGlnValSerTyrThrSerTrpPheLeu-----AspAlaPheAs 294
      |||||
Db 881 AGATGAAATCCCTTTGGCTCCAGGAGATGGATTGGCCCGCGAAAGATCGAAAGG 940
      |||||
QY 294 nTyrAlaIleLeuLysIle-----AspValLeuAsnLeuSerIleGlyGlyProAspPh 313
      |||||
Db 941 AAAAGCGCACCTGAAAGGCTCCCGATGTCGTCACAACTCTTGGCGAGGGAACGCGG 1000
      |||||
QY 313 eMetAspHisProPheValaspLysValTrpGluLeuThrAlaAsnValIleMetVa 333
      |||||
Db 1001 ACTCGATGAATGTTATTTGATATTGTCAAAGCTTGGCGGCGCATGATCTTCCCTGC 1060
      |||||
QY 333 lSerAlaIleGlyAsn-----AspGlyProLeuTyrGlyThrLeuAs 347
      |||||
Db 1061 ATTGACGGCGCAATGTCAGCGAATTCGAACGCGCGACCG-----GGTTCAGTTGA 1114
      |||||
QY 347 nAsnProAlaAspGlnMetAspValIleGlyValGlyGlyIleAspPheGluAspAsnI 367
      |||||
Db 1115 AAACCGCTCACTATCCGCAATCTTCGGATGCGCGATGACAGCTCAACCGCGTT 1174
      |||||
QY 367 eAlaArgPheSerSerArgGlyMetThrTrpGluLeuProGlyGlyTyrGlyArgMe 387
      |||||
Db 1175 AGCGGACTTTTCAATCAAGG-----CCITCGCCATATGACGAAAT 1216
      |||||
QY 387 tLysProAspIleValThrTyrGlyAlaGlyValArgGlySerGlyValLysGlyGlyCy 407
      |||||
Db 1217 CAAGCCGATATTTCTGCACCGCGCTCAGCATCCGCTCTGCTATTCGGGACATAAGTA 1276
      |||||
QY 407 sArgAlaLeuSerGlyThrSerValAlaSerProValValAlaGlyAlaValThrLeuLe 427
      |||||
Db 1277 TGCTGAATGAACGGACATCGATCGCGACCGCATGTATCGGGCATGCTGCCCTTAAT 1336
      |||||
QY 427 uValSerThrValGlnLysArgGluLeuValAsnPro-----AlaSerMetLys 443
      |||||
Db 1337 G-----AGGAA-----GCCAATCGGACCTCACAGTCGATGAATCGA 1375
      |||||
QY 443 sGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMetPheGluGln----- 461
      |||||
Db 1376 GGCATCTCTGCTGAAACCGCGCGCTCTGCTAGTCAAAACCGTTTAAAAAATCCCGGAA 1435
      |||||
QY 462 -----GlyHisGlyLysLeuAspLeuLeuArgAlaTyr 472
      |||||
Db 1436 CAACGCTACGGATACGGTTGGTCAACGCTTAAAGCTTTT 1478
      |||||

RESULT 20
US-09-783-590-349
; Sequence 349, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
```

```
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 349
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (51)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (82)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (205)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (206)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (210)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (230)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (231)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (263)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (298)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (308)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (309)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (310)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (314)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (349)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (355)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (368)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (374)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (380)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (381)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
```

LOCATION: (411)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (421)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (443)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (447)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (448)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (455)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (472)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (477)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (480)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (486)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (489)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-349

Alignment Scores:
Pred. No.: 3,81e-17 Length: 497
Score: 247.00 Matches: 50
Percent Similarity: 76.47% Conservative: 2
Best Local Similarity: 73.53% Mismatches: 8
Query Match: 4.40% Indels: 8
DB: 10 Gaps: 2

US-09-830-837-6 (1-1052) x US-09-783-590-349 (1-497)

QY 735 ValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGlyGlyAlaAsn 754
DB 2 ATTCGGCAGTACACACTTCCAGCAGTGGTGGATCCGGATACCGNAGGAGCTAAC 61
QY 755 IleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyr 774
DB 62 ATCCAGCTCTGAATGAGCTCTGTGTGGAACTGGGTTCAGCAATGGGCTCAGTAT 121
QY 775 GluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAla 794
DB 122 GAAGGGAGTTACCTCGGCCAACCATGAC-----AGTAAGGCT 160
QY 795 LysPheProGluAspGlyValVal 802
DB 161 CTGTTTCTGTAG---GGCGTGGTC 181

RESULT 21

US-09-974-300-1934
Sequence 1934, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Id Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085-500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIORITY FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/660,598

PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1934
LENGTH: 1329
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Alignment Scores:
Pred. No.: 7,04e-16 Length: 1329
Score: 241.50 Matches: 89
Percent Similarity: 45.19% Conservative: 52
Best Local Similarity: 28.53% Mismatches: 108
Query Match: 4.30% Indels: 64
DB: 10 Gaps: 13

US-09-830-837-6 (1-1052) x US-09-974-300-1934 (1-1329)

QY 179 ArgHisSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGln 198
DB 368 CCTCTCTCGACGTAGCGGTGAATCAAGCCATGCCAAAGAGTGTCTAGAAACAAT-CAG 426
QY 199 AlaAspValLeuTrpGlnMetClyThrGlyAlaAsnValArgValAlaValPheAsp 218
DB 427 ACG-----CTGACAGGCAAGAGGAGTGCACAGTCGCTGTCATTGAT 465
QY 219 ThrGlyLeuSerGlyLysHisProHisPheLys-----AsnValLysGluArgThr 235
DB 466 ACGGCGCTA-----TACCTCAGAGATCTTGAAGGAGGATCAGGCGCTTTTCAA 516
QY 236 AsnTrpThrAsnGluArgThr-----LeuAspAspClyLeuGlyHisGlyThrPheVal 253
DB 517 GACTTATTCACACAGAGAACACCTATGATGACAAT---GGCAGCGGCACACACTGC 573
QY 254 AlaGly-----ValIleAlaSerMetArgGluCysGlnGlyPheAlaPro 268
DB 574 GCGGTGATGCTTCGCGAAACGGAGCGCTCATCGGGTCAGTACCGCGGACCTGCTCT 633
QY 269 AspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrp 288
DB 634 GAAGCAGAACTTCGCGTGTAAAGATATTGGACAAAATGGATCCGGATCCGCAACC 693
QY 289 PheLeuAspAlaPheAsnTyrAlaIleLeu-----LysLysIle 301
DB 694 GTCAATCAAGCGCTAGATTGGTCATTCAATTCAATAAGGAAAATCTCTGATGATCCGATC 753
QY 302 AspValLeuAsnLeuSerIleGlyGlyProAspPheMet-----AspHis 316
DB 754 GACATATTTCAATGTCATTTGGTGCAGAGCCTTGGCTACAGAGATGAAGAAGAT 813
QY 317 ProPheValAspLysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIle 336
DB 814 CCAGTCGTTAANGCTGTTCATGCAGCATGGGACCGAGGATGTTGTGTATGCGGCAGCC 873
QY 337 GlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIle 356
DB 874 GGCAACTCCGGTCTCTGATCGCAACGATTCGACCGCGGGTGTCTACAGCAGAGATATT 933
QY 357 GlyValGlyGlyIleAsp-----PheGluAspAsnIleAlaArg 369
DB 934 ACAGTCGGAGCGCTTGGATGACAGGGATACAGTCACCGCGGAGGATGACGATGCGCTCT 993
QY 370 PheSerSerArgGlyMetThrTrpGluLeuProGlyGlyTyrGlyArgMetLysPro 389
DB 994 TATTCAGCAGAGCGCGCAATC-----TATGGTCAAGTCAAAACCG 1035
QY 390 AspIleValThrTyrGlyAlaGlyVal-----ArgGlySerGly----- 402
DB 1036 GACTTGTGGTGGTACCGGCACAAATATTACGTCTGCTTCCACCGCATCTTTCTCGAT 1095

```
QY 403 -----VallySGlyGlyCysArgAlaLeuSerGlyThrSerVal 415
Db 1096 AAGCTGCAAAAACAAACAGAGTCGGCAAAATATATGACATTTGTCGGAACCTCGATG 1155
QY 416 AlaSerProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGlu 435
Db 1156 GCTACGCCGCTGTCGCGAGGAATTCGCCGACATTATC-----CTTCAGCAAGCCCG 1206
QY 436 LeuValAsnProAlaSerMetLysGlnAlaLeuIle 447
Db 1207 GCACAGAACCTGATGATCAAGTCAACAGCTGCTAATG 1242

RESULT 22
US-09-966-921A-5
; Sequence 5, Application US/09966921A
; Patent No. US20020103100A1
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Lassen, Soren
; APPLICANT: Pedersen, Poul
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides Having Proteolytic Activity
; FILE REFERENCE: 10097-200-US
; CURRENT APPLICATION NUMBER: US/09/966, 921A
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1330
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-966-921A-5

Alignment Scores:
Pred. No.: 1,37e-14 Length: 1330
Score: 229.50 Matches: 109
Percent Similarity: 39.46% Conservative: 65
Best Local Similarity: 24.72% Mismatches: 160
Query Match: 4.09% Indels: 107
DB: 10 Gaps: 19

US-09-830-837-6 (1-1052) x US-09-966-921A-5 (1-1330)
QY 57 IleValAlaPheAsnGlyThrPheThrAlaLysAlaArgAsnSerPheIleSerSerAla 76
Db 22 ATTGTGCGAAGCAGCCGACTACTCATTTCTGTGCTTTAGTTCATCGATCGATCGGCT 81
QY 77 -----LeuLysSerSerGluValAlaAspAsnTrpArgIleIleProArg 90
Db 82 GCACCTGCAAAAGACAAAGTTGAGGTAAGGAACAAGATTCATATCGTGTGCTAATCAAA 141
QY 91 -----AsnAsnProSerSerAspTrpProSerAspPheGluVal----- 103
Db 142 GCACCACTACATCAATCAGTACTTTTCAATCAATAGATGATGTCGGTTGGGATTTGGC 201
QY 104 -----IleGlnIleLysGluLysGlnLysAlaGlyLeuThrLeuGlu 118
Db 202 AAAGAGGGATTACACAGATGTTGATGCCAACAG-----CTCCAAACGCTTCAA 252
QY 119 AspHisProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTrp 138
Db 253 AGCAACAAAGACATTCAAATTCAGAAAGGTAATGAATGACAGTAGAAGACTGTTACAACA 312
QY 139 AlaGluSerAspProThr-----ValProCysAsnGluThrArgTrpSerGlnTrpGln 157
Db 313 GAAAAGCGGGAAGTACGCGCGTACCAAGTACACAAACCCCTTGGGGCATAAAG----- 366
QY 158 SerSerArgProLeuArgAlaSerLeuSerLeuLysSerGlyPheTrpHisAlaThr 177
Db 367 -----TCAATTTATATGATCAATCAATTCATAAACAACACT 402
QY 178 GlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeu 197
```

```
Db 403 GGA----- 405
QY 198 GlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPhe 217
Db 406 -----GGCAGCGGAATTAAGTAGCTGTTTAA 432
QY 218 AspThrGlyLeuSerGlnLysHisProHisPheLysAsnValLysGluArg---ThrAsn 236
Db 433 GATACAGGGGTTTATACAGCCATTAGATTAGCTGGTCTGCCGAGCAATGCAAGGAT 492
QY 237 TrpThrAsnGluArgThrLeuAspAspGly-----LeuGlyHisGlyThr 251
Db 493 TTTACCCCAATCTAATCCTTTAGTAGATGTTTCATGCACCGCATGCCCAAGGCGATGTACA 552
QY 252 PheValAlaGlyValIle-----AlaSerMetArgLysGlnGlyPhe 266
Db 553 CATGTTCGCCGAAGTGTATTGGCGCATGGAGCGAGTAATGGACAGCGCTTACGGGGTG 612
QY 267 AlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTrpThr 286
Db 613 GTCGCCAAGCCAACTATGGCATATAAAGTATTAGGAGATAACGCCAGCGGATCTCT 672
QY 287 SerTrpPheLeuAspAlaPheAsnTrpAlaIleLeuLysLysIleAsp----- 302
Db 673 -----GATGATATTGCAGCAGCTATCAGACATGTAGCTCATGAAGCTTCACGT 720
QY 303 -----ValLeuAsnLeuSerIleGly-----GlyProAspPheMetAsp 315
Db 721 ACAGGTTCCAAAGTAGTAATTAATATGTCGTAGTTCATCTGCCAAGGATCATTCATGAT 780
QY 316 HisProPheValAspLysValTrpGluLeuThrAlaAsnValIleMetValSerAla 335
Db 781 GCTAGTGCAGTAGATTATGCATAT-----GGAAAAGGCTATTATCTGCTGCGG 831
QY 336 IleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspVal 355
Db 832 GCTGGTAAATAGTGGGTGAGGAGCAATAACATCGGCTTCTCGCGGGCTTGTAAATGCA 891
QY 356 IleGlyValGlyGlyIleAsp---PheGluAspAsn-----IleAlaArgPhe 370
Db 892 GTGGCAGTAGCGCGCATTTGGAGAAATGTCAGCAAAATGGAACTATTCGATGCTGATTC 951
QY 371 SerSerArgGlyMetThrTrpGluLeuProGlyGlyTyrGlyArgMetLysProAsp 390
Db 952 TCATCTAGAGGAATCCGCA-----ACTGCTGGAGATTATATCATTCATCAAGAGCGTGAT 1005
QY 391 IleValThrTyr-----GlyAlaGlyValArgGlySerGlyValLysGlyCysArg 408
Db 1006 ATTGAAGTTTCAGCTCCGGGAGCAGTGTAAGTCTACATGTGTACACTGCGCGTTATAAT 1065
QY 409 AlaLeuSerGlyThrSerValAlaSerProValValAlaGly----- 422
Db 1066 ACGATCAGCGGTACATCAATGGCTACACCTCATAGCTAGCTGGTGTAGCTGCTAAAATCTGG 1125
QY 423 ---AlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSer 441
Db 1126 TCAGCGAATACCTTCATTAAAGTCAATGCCCAACTGCCGACAGAAATGCCGCTTAA 1185
QY 442 Met 442
Db 1186 GTA 1188

RESULT 23
US-09-966-921A-1
; Sequence 1, Application US/09966921A
; Patent No. US20020103100A1
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Lassen, Soren
; APPLICANT: Pedersen, Poul
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides Having Proteolytic Activity
; FILE REFERENCE: 10097-200-US
```

;; CURRENT APPLICATION NUMBER: US/09/966,921A
;; CURRENT FILING DATE: 2001-09-28
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 1

;; LENGTH: 1306
;; TYPE: DNA
;; ORGANISM: Bacillus sp.
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (50)..(1303)
;; OTHER INFORMATION:
;; NAME/KEY: mat_peptide
;; LOCATION: (371)..()
;; OTHER INFORMATION:
US-09-966-921A-1

Alignment Scores:
Pred. No.: 1,92e-14 Length: 1306
Score: 228.00 Matches: 114
Percent Similarity: 38.76% Conservative: 67
Best local Similarity: 24.41% Mismatches: 161
Query Match: 4.06% Indels: 125
Gaps: 21

US-09-830-837-6 (1-1052) x US-09-966-921A-1 (1-1306)

QY 26 GluLysSerPheGluLysAlaProCysProGlyCysSerHisLeuThrLeuLysVal 45
|||||
DB 55 GAAAGAGAGCATT-----TGCAGCCACATTACT----- 84
QY 46 GluPheSerThrValValGluLysGluThrValAlaPheAsnGlyTyrPheTh 65
|||||
DB 85 -----CAGTATTACGATGGGATTATCCGTATTTCACAGGAGCACT 126
QY 65 rAlaLysAlaArgAsnSerPheLeuSerAlaLeuLysSerSerGluValAlaAsnTr 85
|||||
DB 127 TCAAAAGACAAA-----GTTGAGGTGAAGGACAGATTCATA 165
QY 85 pArgIleIleProArg-----AsnAsnProSerAspTyrProSerAspPheGluVa 103
|||||
DB 166 TCGTGTGTAATCAAGACCACTACATCAATCACTATTTCATCAATCAATAGATGT 225
QY 103 l-----lleGlnIleLysGluLysGlnLysAlaGl 113
DB 226 CGGTGGGANTTTGGCAAGAGGGNTTTACAACAGATGTGTGATGCCAAACAG----- 277
QY 113 yLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysValPh 133
DB 278 -CTCAACGCTTCAAGCAACAAAGCATTCATTAATTCAGAGTAATCAATGACAGT 336
QY 133 eArgSerLeuLysTyrAlaGluSerAspProThr-----ValProCysAsnGluThrArgTr 152
DB 337 AGAACTGTTACACAGAAAGGGGAGTACGCGGTACCAAGTACACAAACCCCTTG 396
QY 152 pSerGlnLysTrpGlnSerSerArgProLeuArgAlaLaserLeuSerLeuGlySerGl 172
DB 397 GGCATAAAG-----TCAATTTATTAATGATCAATC 426
QY 172 yPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArgGl 192
DB 427 AATTCAAAAACTGGA----- 445
QY 192 nValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnVa 212
DB 446 -----GGCAGCGGAAT 456
QY 212 lArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsnVally 232
DB 457 TAAGTACTGCTTTTAGATACAGGGGTTTATACAAAGCCATTAGATTACTGGTCTGC 516
QY 232 sGluArg---ThrAsnTrpThrAsnGluArgThrLeuAspAspGly----- 246
|||||

DB 517 CGACAAATGCAAGGATTTTACCCTAATCTTCTAGTAGATGGTTTCATGCACCGATCG 576
QY 247 -LeuGlyHisGlyThrPheValAlaGlyValIle-----AlaSerMetar 261
DB 577 CCAAGGCGATGGTACACATGTTCCGGAATGTTGGCGCATGGAGCGAGTAAATGACA 636
QY 261 gLUCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAs 281
DB 637 AGCGTTTACGGGTGGCTCCGCAAGCAACTGGCATATATAAGTATTAGAGATAA 696
QY 281 nGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIl 301
DB 697 CGCAGCGGATCTCT-----GATGATATTGCAGCAGCATATCAGACATGAGC 744
QY 301 eAsp-----ValLeuAsnLeuSerIleGly-----G1 310
DB 745 TGATGAAGCTTTCACGTACAGGTTCCAAAGTAGTAATTAATATGCTAGGTTTCATCTGC 804
QY 310 yProAspPheMetAspHisProPheValAspLysValTrpGluLeuThrAlaAsnAsnVa 330
DB 805 CAAGGATTCATTGATTGCTAGTCAGTAGATTATGCAAT-----GGAAAAGGTGT 855
QY 330 lIleMetValSerAlaIleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAl 350
DB 856 ATTAAATCGTTGCTCGGCTGTTAATAGTGGTCAGGCAACAATCAATCGCTTCTCTGG 915
QY 350 aAspGlnMetAspValIleGlyValGlyIleAsp-----PheGluAspAsn----- 366
DB 916 CGGCTGTGAATATGCAATGCGCAGTAGCGGCATGGAGATGTTACGAAATGGAACCTTA 975
QY 367 -----lIleAlaArgPheSerSerArgGlyMetThrTrpGluLeuProGlyGlyTyrGl 385
DB 976 TCGAGTAGCTAGTATTCTCATCTAGAGGAATCCGCA-----ACTGCTGGAGATTATAT 1029
QY 385 yArgMetLysProAspIleValThrTyr-----GlyAlaGlyValArgGlySerGlyVa 403
DB 1030 CATTCAAGACGCGTATTAAGTTTACGTCGGGAGCAAGTAGTAGTCTCATCGTA 1089
QY 403 lLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValAlaIleGly-- 422
DB 1090 CACTGGCGGTATAATACCATCAGCGGTACATCAATGCTACACCTCATGTAGTGGGTT 1149
QY 423 -----AlaValThrLeuValSerThrValGlnLysArgGluLe 436
DB 1150 AGCTGCTAAAATCTGGTCAGCGAATACTTCAATTAAGTCATAGCAACTGCGCAGAGATT 1209
QY 436 uValAsnProAlaSerMet 442
DB 1210 GCATAATCGCGCTAAAGTA 1228
RESULT 24
US-10-209-812-1
;; Sequence 1, Application US/10209812
;; Publication No. US20030087785A1
;; GENERAL INFORMATION:
;; APPLICANT: Olsen, Arne Agerlin
;; APPLICANT: Osten, Claus von der
;; APPLICANT: Andersen, Kim Villbour
;; APPLICANT: Ernst, Steffen
;; APPLICANT: Roggen, Erwin Ludo
;; TITLE OF INVENTION: A Modified Polypeptide
;; FILE REFERENCE: 5666.200-US
;; CURRENT APPLICATION NUMBER: US/10/209,812
;; CURRENT FILING DATE: 2002-08-01
;; PRIOR APPLICATION NUMBER: US/09/417,359
;; PRIOR FILING DATE: 1999-10-13
;; PRIOR APPLICATION NUMBER: PA 1998 01301
;; PRIOR FILING DATE: 1998-10-13
;; PRIOR APPLICATION NUMBER: PA 1999 01418
;; PRIOR FILING DATE: 1999-10-04
;; PRIOR APPLICATION NUMBER: 60/105,624
;; PRIOR FILING DATE: 1998-10-26
;; PRIOR APPLICATION NUMBER: 60/157,426

;; PRIOR FILING DATE: 1999-10-04

;; NUMBER OF SEQ ID NOS: 6

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 1

;; LENGTH: 840

;; TYPE: DNA

;; ORGANISM: Bacillus sp.

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: (1)...(840)

US-10-209-812-1

Alignment Scores:

Pred. No.: 1,08e-14 Length: 840

Score: 227.50 Matches: 85

Percent Similarity: 41.24% Conservative: 35

Best Local Similarity: 29.21% Mismatches: 128

Query Match: 4.05% Indels: 43

DB: 9 Gaps: 13

US-09-830-837-6 (1-1052) x US-10-209-812-1 (1-840)

QY 195 GlnThrLeuGlnAlaLeuValLeuTrpGlnMetGlyThrGlyAlaAsnValArgVal 214

DB 46 CAAACACCTCAACCCCTGCTGGGATGATTAACCCGGAAGCAGCACTCAAAACGGTG 105

QY 215 AlaValPheAspThrGlyLeuSerGluLysHisProHisPhe-----LysAsnValLys 232

DB 106 GCGGCTTGAATCCGGATGGATTAACCAACCCCTGATCTTGCAAGAAAGTAATAAAA 165

QY 233 -----GluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGlyLeu 247

DB 166 GGTACGAGCTTATCGACAGGGACAT-----AACCAATGATCTTAAC----- 210

QY 248 GlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263

DB 211 GGACATGGTACCATCTGTCGGTACTGTGCTGATACGAACAATGGAATGGCGTA 270

QY 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal 283

DB 271 GCGGTATGACACAGATACGAGATCCCTGCGGTACGGGTCTTGTATGCCAATGAAGT 330

QY 284 SerTyThrSerTrpPheLeuAspAlaPheAsnTyAlaIleLeuLysLysIleAspVal 303

DB 331 GGCTACTTGACAGCATTCCTCAGTATCCGCTATGCTGCTATCAAGGGCAAGGTA 390

QY 304 LeuAsnLeuSerIleGlyPro-----AspPheMetAspHisProPheValAspLys 321

DB 391 CTCACCTCTCCCTGCTGCGAATGCAACTCCACAACTCTTAAGAGTCCCGTCGACTAT 450

QY 322 ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341

DB 451 GCATGGAAC-----AAAGAGCTGTAGTCTGTGCTGCGAGGAATGACAAT--- 498

QY 342 LeuTyThrGlyThrLeuAsnProAlaAspGlnMetAspValIleGlyValGlyIle 361

DB 499 ---GTATCCCGTACATTCACACAGCTTCTTACCCTAATGCCATTCAGTAGTGCAT 555

QY 362 AspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrTrpGluLeuPro 381

DB 556 GACTCCAAATGATCGAAAGCATCTCTCCAATACGGA-----ACGTGG----- 600

QY 382 GlyGlyTyArgMetLysProAspIleValThrTyThrGlyAlaGlyValArgGlySer 401

DB 601 -----GTGGATGTCATCTGCTCCAGGTGTGAACATAGATCAATCAACC 639

QY 402 GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValValAla 421

DB 640 GTTCCGAATAATGGCTACTCTACATGCTGTGGTACGTCCATGCCATCCCTCAGTGGCC 699

QY 422 GlyAlaValThrLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSer 441

DB 700 GGTGCTGCTGCTTGTGGCAAGTCAAGGTAAG-----AATAACGTACAA 744

QY 442 MetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMet---PheGlu 460

DB 745 ATCCGCCAGGCCATTGAGCAAAACCCGCGATAGATCTCTGGCACTGGAAACAACTTCAG 804

QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAla 471

DB 805 -----TATGGTAAATCAACTCAAAACAAAGCT 831

RESULT 25

US-09-974-300-1933

;; Sequence 1933, Application US/09974300

;; Patent No. US20020146721A1

;; GENERAL INFORMATION:

;; APPLICANT: Berk, Randy M.

;; TITLE OF INVENTION: Methods for Monitoring Multiple Gene

;; FILE REFERENCE: 10085.500-US

;; CURRENT APPLICATION NUMBER: US/09/974,300

;; PRIOR FILING DATE: 2001-10-05

;; PRIOR APPLICATION NUMBER: 09/680,598

;; PRIOR FILING DATE: 2000-10-06

;; PRIOR APPLICATION NUMBER: 60/279,526

;; NUMBER OF SEQ ID NOS: 8481

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 1933

;; LENGTH: 1397

;; TYPE: DNA

;; ORGANISM: Bacillus licheniformis

US-09-974-300-1933

Alignment Scores:

Pred. No.: 3.97e-14 Length: 1397

Score: 225.50 Matches: 108

Percent Similarity: 41.22% Conservative: 61

Best Local Similarity: 26.34% Mismatches: 173

Query Match: 4.01% Indels: 71

DB: 10 Gaps: 17

US-09-830-837-6 (1-1052) x US-09-974-300-1933 (1-1397)

QY 50 ThrValValGluTyTrpGluTyIleValAlaPheAsnGlyTyTrpPheThrAlaLysAlaArg 69

DB 10 AGTATCGTGGCTAT---TTTGTATGGCTTTTATCTATATT---GCCTTA 57

QY 70 AsnSerPheIleSerSerAla-----LeuLysSerSerGluVal 82

DB 58 TCCACATCTCTACCCGAGTGCGAGCAACTTCGTTCCCGATATAAAGTCGCTGAGCTT 117

QY 83 AspAsnTrpArgIleIleProArgAsnAsnProSerSerAspTyTrpSerAspPheGlu 102

DB 118 GAGAAAGCTGAAATCTACCGTGTATATGATGTGACGCTCTGATAAACAACAGCAGCTT--- 174

QY 103 ValIleGlnTleLysGluLysGlnLysAla---GlyLeuLeuThrLeuGluAspHisPro 121

DB 175 ATCGTGGAACTGAAAGAAAGTCCGTTGCCAAGCAAAACGCGGAGGAGAAACAAACAG 234

QY 122 AsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyAlaGluSer 141

DB 235 AAGGCTTCTCTAAACACAGCTCGAAGCAAAAGCGCTGAAACACACTTAAAAA-AGCGAAGT 293

QY 142 AspProThrVal---ProCys-----AsnGluThrArgTrpSerGlnLys 155

DB 294 AAACCCGAATACGACCGTGTATTTTCGGCTTTTCTATGAAACTCCGCCCAAGT-AAA 352

QY 156 TrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySerGlyPheTrpHis 175

DB 353 TTCCAAGCTGCTCGCCGTCAGGAAGTTAAAGCGGTTTATCCAAACGCAACTTACAAAC 412

QY 176 AlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIlePro----- 190

Db 413 CTGACGCGTAAAGAAAGACGCTGACACTTCGACGACGCGCATTTATCCACAGATGG 472
QY 191 ArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrGlyAla 210
Db 473 ATAAAGTGCCTTC-ATCGGACGGATCAGGATGGAATCGGCTATACCGGTAAA 531
QY 211 AsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsn 230
Db 532 GGCATTAGGTAGCGTATTGATACGGGTGATTACACCCATCCGACTTAAGAAT 591
QY 231 -----ValLysGluArgThrAsnTrpThrAsnGluArg 241
Db 592 AATTCGCGCCATATAAGGCGTATGACTTTGTGTAATATTATGATCCACAGGAACA 651
QY 242 ThrLeuAspAspGlyLeuGly-----HisGlyThrPheValAlaGlyValIle 257
Db 652 CCGACAGCGACCCGCGCGGAGCGACTGATCAGCGCACTCATGTTGCAGGAACAATC 711
QY 258 AlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgVal 277
Db 712 GCCGGAACGACAGATTAAAGCGGTGGCGCTGAAGCAACGCTTCTTGCTACCGCGTG 771
QY 278 PheThrAsnAsnGlnValSerTyThrSerTrpPheLeuAspAlaPheAsnTyralIle 297
Db 772 CTCGCGCCCGCGGCTCAGGCACACCAAGATGCTGTCGGGCATTTGAAAGCCGTT 831
QY 298 LeuLysLysIleAspValLeuAsnLeuSerIleGly-----GlyProaspPhe 313
Db 832 GCAGACGGAGCAAAAGTGATGAACCTCTCATTTGGAAATTCCTGAAACAGCCCTGATTAT 891
QY 314 MetAspHisProPheValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetVal 333
Db 892 GCAACAAGCATCGCACTGGAC-----TGGCGCATG--GCTGAAGGGGTGTGCGCGTT 942
QY 334 SerAlaIleGlyAsnAspGlyProLeuTyrglyThrLeuAsnAsnProAlaAspGlnMet 353
Db 943 AGCTAAACGGTAAACAGCGACCGGAAACTGGACGCTCGCTCGCGGGAACATCAAGG 1002
QY 354 AspValIleGlyValGlyIle-----AspPheGluAspAsnIleAla 368
Db 1003 GTTGCATTTCCGTCGCGCATCACAGCTTCGCTATATAGTATTCGGTGACGCTTCT 1062
QY 369 ArgPheSerSerArgGlyMetThrThrTrpGluLeuProGlyGlyTyrglyArgMetLys 388
Db 1063 TCGTATTTCATCAGCAAGGTGATG-----GGATACCAAGAGGAAAG 1104
QY 389 Pro-----AspIleValThrTyrglyValGly----- 397
Db 1105 GATCTTGAAGCATTAACGGTCAACAGGTGGAACTGTTGAAGCGGCGCTCGCACAGCT 1164
QY 398 -----ValArgGlySerGlyValIleGly 405
Db 1165 GACGATTCTCAGGCAAGATGTGAAGGA 1194

Search completed: June 2, 2003, 06:02:21
Job time : 1152 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 2, 2003, 02:29:02 ; Search time 6516 seconds
(without alignments)
4698.611 Million cell updates/sec

Title: US-09-830-837-6
Perfect score: 5617
Sequence: 1 MKLVNTWLLVLLCGKKH.....PRVKRQLMQVHPKTPSV 1052

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09830837/runat_23052003_181924_8065/app_query.fasta_1.1223
-DB=GenEmbl -QFWT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=ptt -THR_MAX=100 -THR_MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=ptt -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830837.ecgn_1.1.4449 @runat_23052003_181924_8065 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

Result No.	Score	Query Match %	Length	ID
1	5617	100.0	4338	6 AX024613
2	5617	100.0	4338	9 HUMKIAAL
3	5487	97.7	4198	10 AF078105
4	5476	97.5	3895	6 AX024609
5	5476	97.5	3895	10 AF094821
6	5427	96.6	3788	6 AX024611
7	5427	96.6	3788	10 AF094820
8	5412.5	96.4	4204	10 BC011533
9	3069	54.6	3026	9 BC026330
10	3063	54.5	3085	9 AK091212
11	2383.5	42.4	3138	3 AF441758
12	2219.5	39.5	174920	3 AC010701
13	2219.5	39.5	294308	3 AE003595
14	1965.5	35.0	100592	8 T29J13
15	1772	31.5	115884	2 AC017581
16	1716	30.6	1650	9 HSM801435
17	1355	24.1	170371	8 AB026295
18	1208.5	21.5	105815	8 AP002536
19	850.5	15.1	113320	2 AC010662
20	812	14.5	3333	9 AK054744
21	774.5	13.8	129955	2 AC010056
22	774.5	13.8	129955	2 AC011909
23	728	13.0	214287	2 AC040169
24	567	10.1	2905	9 AB047598
25	544.5	9.7	195287	2 AC112476
26	443.5	7.9	252539	2 AC098090
27	381.5	6.8	32704	1 SC8A11
28	373.5	6.6	42527	1 SC51A
29	363.5	6.5	3348	1 D83672
30	352.5	6.3	36102	3 LMFLA325
31	342	6.1	2494	1 BACAPRJ
32	341.5	6.1	1527	1 BACSAS
33	341	6.1	1220	6 I01972
34	341	6.1	1220	6 I08532
35	341	6.1	1524	6 I03244
36	339	6.0	1473	1 BACSUBTI
37	339	6.0	1473	1 S51909
38	339	6.0	2044	6 E01654
39	339	6.0	2044	6 E02803
40	338.5	6.0	1146	1 AF368283
41	338.5	6.0	1500	1 BACSETL
42	338.5	6.0	1500	6 E03694
43	338.5	6.0	1500	6 I05124
44	338.5	6.0	1500	6 I07865
45	338.5	6.0	1500	6 I95503

ALIGNMENTS

RESULT 1

QY	441	SerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGlu	460
DB	1817	AGTATGAACAGCGCCCTGATCGGTCAGCCCGAGGCTCCCGGGGTCAACATGTTGAG	1876
QY	461	GlnGlyHisGlyLysLeuAspLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro	480
DB	1877	CAAGCCAGCGGAAGCTCATCTGCTCAGAGCTATCAGATCCTCAACAGCTACAAGCCA	1936
QY	481	GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr	500
DB	1937	CAGGCAAGTTTGAGCCCAAGCTACATAGATCGACTGCTCCCTACATGTGGCCCTAC	1996
QY	501	CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn	520
DB	1997	TGCTCCCAAGCCATCTACTATGGAGGAATGCCAGACATGTGTAAATGTCACCATCTCAAC	2056
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln	540
DB	2057	GGCATGGGAGTCACAGGAAGAATTGTAGATAAGCCTGACTGGCAGCCCTATTGTGCCACAG	2116
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTrpSerGly	560
DB	2117	AACGGAGACAACATTGAAGTTCCCTCTCTACTCTCTCGGTCTTATGGCTTTGGTGGGC	2176
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln	580
DB	2177	TACCTGGCCATCTCCATTTCTGTGACCAAGAAGCGGCTTCTGGGAAGCAATGCTCAG	2236
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600
DB	2237	GGCCATGTGATCATCTGCTGCTTCCCGAGAGACAGACAGATCAAAAAATGGTGCAGAA	2296
QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProArgSer	620
DB	2297	CAGACTTCACAGTAAAGCTCCCAATTAAGGTGAAGATAATTCCTACTCCCGCGGAAGC	2356
QY	621	LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg	640
DB	2357	AAGAGAGTTCTCTGGGATCAGTACCACACCTCCCGTATCCACCTGGCTATTTCGCCAGG	2416
QY	641	AspAsnLeuArgMetLysAsnProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660
DB	2417	GATAATTTAAGATGAAGATCAACCTTTAGACTGGAATGGTATGCATCCACATCCACCAAT	2476
QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
DB	2477	TTCAGGGATATGTACAGCATCTGAGAAGCATGGGCTACTTTGTAGAGTCTCTCGGGGCC	2536
QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
DB	2537	CCCTTCACGTGTTTGTATGCCAGTCAGTATGGCACTTTGCTGATGGTGACAGTGAAGAG	2596
QY	701	GluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu	720
DB	2597	GAGTACTTCCCTGAAGAGATCGCAAGCTCCCGAGGGAGCTGGACAACGGCTCTCGCTC	2656
QY	721	ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
DB	2657	GTCATCTTCAGTGACTGGTACACACTCTCTGTATGAGAAAAGTGAAGTTTATGATGAA	2716
QY	741	AsnThrArgGlnTrpTrpMetProAspThrGlyValAlaAsnIleProAlaLeuAsnGlu	760
DB	2717	NACACAAGCACTGGTGGATCGGATACCGGAGGAGTAAATCCCACTCTGAATGAG	2776
QY	761	LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu	780
DB	2777	CTGCTCTCTGTGGAACATGGGTTTCAGCGATGGCTGTATGAAGGGGAGTTACCCCTG	2836
QY	781	AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800
DB	2837	GCCAAACCATGACATGATTATCGTCAGGGTCAGCATCCGAAGTTTCCAGAAGATGCG	2896
QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla	820
DB	2897	GTCGTGATPAACACAGACTTTCAAGCAAGGATGGAGGTTTAAAGCAGGAACAGCA	2956
QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
DB	2957	GTTGTTGAAAGCTCCCACTTTTGGGACTTTATCAGATTCAGATTCAGGAGTGGAGCCGG	3016
QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
DB	3017	ATTGTACTATATGGGACTCCAATTGCTTGGATGACAGTCACCGACAGAGGACTGCTTT	3076
QY	861	TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProSerLeuSer	880
DB	3077	TGGCTTCTGGATGCTCTCTCCAGTACATCGTATGGGTGACACCGCTAGCCTCAGT	3136
QY	881	HisSerGlyAsnArgGlnArgProSerGlyAlaGlySerValThrProGluArgMet	900
DB	3137	CACCTCGGAACCGCCAGCGCTCTCCAGTGGAGCAGGCTCAGTCACTCCAGAGAGATG	3196
QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
DB	3197	GAAGGAACCATCTTCATCGGTACTCCAAGGTTCTGGAGGCCCATTTGGGAGACCCAAA	3256
QY	921	ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu	940
DB	3257	CCTCGGCTCTACCGCTTGTCCAGCTTGTCTTGGGCCAAGCCACAGCCTTTAAACGAG	3316
QY	941	ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal	960
DB	3317	ACGGGCCCACTAAACCTTTTGGAAACATCAGAACTACTCTCCATTGACCTGGACAAGGTG	3376
QY	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer	980
DB	3377	GTTTATCCCACTTTCATCGAATCCCTCAAGTGAAGCCCTTGTCCCTGGAGAGAGC	3436
QY	981	GlyAlaTrpAspIleProGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln	1000
DB	3437	GGCGCTGGACATCTCTGGAGGATCATCGCTGGCGCTACAAACAGAGGTTGGGCCAG	3496
QY	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln	1020
DB	3497	ACCATTCTCTTCTTCCCTTCTGGAGCCATGTGTGCTCTGGCCCTCTTTGTGGTACAA	3556
QY	1021	IleAsnLysAlaLysSerArgProLysArgAlaGlyLysProArgValLysArgProGlnLeu	1040
DB	3557	ATCAACAGCCCAAGACAGAGCGCGAAGCGGAGGCCAGGCTGAAGCGCCGCGCAGCTC	3616
QY	1041	MetGlnGlnValHisProLysThrProSerVal	1052
DB	3617	ATGCAGCAGTTTCAACCGCCCAAGACCCCTTCGGTG	3652
RESULT 2			
HUMKIAL			
LOCUS			
DEFINITION			
HUMAN mRNA for KIAA0091 gene, complete cds.			
ACCESSION			
D42053			
VERSION			
D42053.1 GI:577308			
KEYWORDS			
KIAA0091.			
SOURCE			
Homo sapiens			
ORGANISM			
Homo sapiens male myeloblast cell_line KG-1 cDNA to mRNA.			
REFERENCE			
AUTHORS			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
TITLE			
Prediction of the coding sequences of unidentified human genes.			
III. The coding sequences of 40 new genes (KIAA0081-KIAA0120)			
deduced by analysis of cDNA clones from human cell line KG-1			
JOURNAL			
DNA Res. 2 (1), 37-43 (1995)			
MEDLINE			
95308325			
REFERENCE			
2 (bases 1 to 4338)			
AUTHORS			
Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.			

QY	461	GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro	480
DB	1877	CAAGGCCACGGCAGCTCGATCTGCTCAGAGCCCTATCAGATCTCAACAGCTACAGCCA	1936
QY	481	GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTyrProTyr	500
DB	1937	CAGGCAAGTTTGAGCCCGCAGCTACATAGATCTCAGTGTGCTCCCTACATGTGGCCCTAC	1996
QY	501	CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn	520
DB	1997	TGCTCCAGCCCAATCTACTANGAGGAATCCGACAGTGTAAATGTCAACATCCCAAC	2056
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTyrGlnProTyrLeuProGln	540
DB	2057	GGCATGGAGTCAACAGAAATGTAGATAAGCCTGACTGGCAGCCCTATTTGCCACAG	2116
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTyrProTyrSerGly	560
DB	2117	AACGGAGACAACATTAAGTTGCTTCTCTCTACTCTCGGTCTTATGGCCTTGGTGGGC	2176
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTyrGluGlyIleAlaGln	580
DB	2177	TACTTGGCCATCTCCATTTCTGTGACCAAGAACGGCTTCTTGGGAAGCATTTGCTCAG	2236
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600
DB	2237	GGCCATGTCTATGATCACTGTGGCTTCCCCAGCAGACAGAGTCAAAAATGTGCAGAA	2296
QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleProThrProProArgSer	620
DB	2297	CAGACTTCACAGTAAGTCCCCATTAAGTGAAGATAATCTCTACTCCCGCGGAAGC	2356
QY	621	LysArgValLeuTyrPaspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg	640
DB	2357	AACAGAGTTCTCTGGGATCAGTACCACACCTCCGCTATCCACTGGCTATTTCCCGAGG	2416
QY	641	AspAsnLeuArgMetLysAsnAspProLeuAspTyrPaspGlyAspHisIleHisThrAsn	660
DB	2417	GATAATTTAAGATGAAGATGACCTTTAGACTGGAATGGTATCAGTATCCACATCCACCAAT	2476
QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
DB	2477	TTCAGGGATATGTACCAGATCTTGAGAGCATGGGCTACTTTGTAGAGTCTCTCGGGGCC	2536
QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
DB	2537	CCCTTCACGTGTTTGTATGCCAGTCAATGATGGACCTTTGCTGATGGTGNACATGAGGAG	2596
QY	701	GluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu	720
DB	2597	GAGTACTTCCCTGAAGAGATCGCAAGCTCCGGAGGACGTGACAAACGGCTCTCGCTC	2656
QY	721	ValIlePheSerAspTyrPyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
DB	2657	GTCACTCTCAGTACGTCACACACTCTGTATGAGAAAAGTGAAGTTTATGATGAA	2716
QY	741	AsnThrArgGlnTyrPaspTyrMetProAspThrGlyValAlaAsnIleProAlaLeuAsnGlu	760
DB	2717	AAACACAGGCGATGGTGGATGCGGATACCGGAGGAGCTAACATCCAGCTCTCAATGAG	2776
QY	761	LeuLeuSerValTyrPaspMetGlyPheSerAspGlyLeuTyrGluGlyPheThrLeu	780
DB	2777	CTGCTCTCTGTGGAACATGGGTTCCAGCATGGCTGTATGAAGGGAGTTTCAACCTG	2836
QY	781	AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800
DB	2837	GCAACCATTCACATGATATTATGCTCAGGGTGCAGCATCGCAAGTTTCCAGAAGATGC	2896
QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla	820
DB	2897	GTCTGTATACACAGACTTTCAAGGACCAAGGATTGGAGTTTAAAGCAGGAAACAGCA	2956
QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
DB	2957	GTGTTGTGAAAACGTCCTCCCATTTTGGGACATTATCAGATCCAGCTGAGGGTGGAGCGCG	3016
QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
DB	3017	ATTGTACTGTATGGGACTCCAATTCGTTGGATGACAGTCCAGCAGACAGAGGACTGCTTT	3076
QY	861	TyrLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer	880
DB	3077	TGGCTTCTGGATGCCCTCTCCAGTACATCATGATGGGTGACACCGCTAGCTCAGT	3136
QY	881	HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet	900
DB	3137	CACCTCTGGGAACGCCAGCGCCCTCCAGTGGAGCAGGCTCAGTCTCCAGAGAGGATG	3196
QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
DB	3197	GAAGGAACCATCTTCATCGGTACTCCAAGTTCTGGAGGCCATTTGGGAGACCCAAAA	3256
QY	921	ProArgProLeuProAlaCysProArgLeuSerTyrAlaLysProGlnProLeuAsnGlu	940
DB	3257	CCTCGGCCTCTACAGCTGTCCAGCTTGTCTGTGGCCCAAGCCACAGCCTTTAAACGAG	3316
QY	941	ThrAlaProSerAsnLeuTyrLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal	960
DB	3317	ACGGCGCCAGTAGTAACTTTGGAAACATCAGAAGCTACTCTCCATTGACCTGGACAAGGTG	3376
QY	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer	980
DB	3377	GTGTTTACCACAACTTTCATCGAATCGCCCTCAAGTAGGAGGCCCTTGTCCCTGGAGAGC	3436
QY	981	GlyAlaTyrPaspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln	1000
DB	3437	GGCGCCTGGGACATCTCTGGAGGATCATGCTTGGCGCTACAAACAGAGAGTGGGCCAG	3496
QY	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValGln	1020
DB	3497	ACCATCTCTGCTTCTTCTTGGGAGCATGGTGTCTGTGCTTCTTGTGTGTACAA	3556
QY	1021	IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu	1040
DB	3557	ATCAACAAAGCCCAAGACAGCGGCGAAGCGGAGGAGCCCAAGGTGAAGCGCCCGCAGCTC	3616
QY	1041	MetGlnGlnValHisProProLysThrProSerVal	1052
DB	3617	ATGCACAGGTTTCAACCGCCCAAGACCCCTTCTCGGTG	3652
RESULT 3			
AF078105		4198 bp mRNA linear	ROD 19-NOV-1998
LOCUS		Cricetulus griseus site-1 protease of sterol regulatory element	
DEFINITION		binding proteins mRNA, complete cds.	
ACCESSION	AF078105		
VERSION	AF078105.1	GI:38922203	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			

Db	1887	TGTTCTCAGCCCATCTACTATGAGGAATGCCAACAAATTGTTAATGTCAACCATCTCAAT	1946
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln	540
Db	1947	GGCATGGAGTCACAGAGAATTGGATAAGCCTGAGTGGCGCCCTATTATTACCACAG	2006
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGly	560
Db	2007	AATGGAGACAACATTGAAGTGGCCTCTCTACTCTCAGTGTTATGGCCCTGGTCAGGC	2066
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln	580
Db	2067	TACCTGGCCATCTCCCATTTCTGTGACCAAGAGCAGCTTCTCTGGGAAGGCATTGCACAG	2126
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrLysSerLysAsnGlyAlaGlu	600
Db	2127	GGTCACATCATGATCAGGTGGCTTCCCCAGCAGACGGAACAAAATGTGTCCGAG	2186
QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProArgSer	620
Db	2187	CATACCTCCACAGTGAAGCTTCCCATTAAGGTGAAGATCATTCCTCCACCCCTCTCTCGGAGC	2246
QY	621	LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg	640
Db	2247	AGAGAGTCTCTGGGACCAAGTATCACACCTCCGCTACCCCGCCAGGCTACTTTTCCAGG	2306
QY	641	AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660
Db	2307	GACAACTTGGGATGAAGATGATCTTTAGACTGGATGGCAGCCATGTCTCCACCAAT	2366
QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
Db	2367	TTCAGGACATGTACCAGCACCTTGCAGCATGGCTACTTCTGTGAGGTGCTCGGTGCC	2426
QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
Db	2427	CCATTCACGTGCTTGTATGATACACAGTATGGCAGCTTGCCTATGTGGATAGTGAAGA	2486
QY	701	GluTyrPheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu	720
Db	2487	GAGTACTTCCAGAGGAGATTGCCAAGCTGAGAGGAGCGTGACAAATGGCCCTTCCCTC	2546
QY	721	ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
Db	2547	GTCACTTTCAGTCAGTGGTACACACTTCTGTTATGAGAAAAGTGAAGTTTACGATGAA	2606
QY	741	AsnThrArgGlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGlu	760
Db	2607	AACAAAGGCGAGTGGTGGATGCCAGATACTGGAGGAGCCCAACATCCAGCTCTGAACGAG	2666
QY	761	LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu	780
Db	2667	CTGCTGCTGTGTGGAACTGGGGTTCAGCGATGGCCTTATGAAAGGGGATTTGCCCTG	2726
QY	781	AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800
Db	2727	GCGAATCATGACATGATTATGATCATCGGATGCAGCATCCCAAGTTTCCAGAAGATGTT	2786
QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla	820
Db	2787	GTGTGATCACACAGACTTTCAGAGCACAAGGATTTGAGGTCTTTAAACAAGACAGCA	2846
QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
Db	2847	GTGTGTAAGAACTTCCCATTTTGGGCTTTATCAGATTCACACTGAAGTGGGGCCGG	2906
QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
Db	2907	ATCGTGTGTATGAGATTCATTCATTGCTGGATGCAGTCACAGACAGAGGATTCGTTT	2966
QY	861	TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer	880
Db	2967	TGGCTTCTGGATCGACTCTTCACTACATCATATATGGGTGAACCTCCAGCCCTCAGC	3026
QY	881	HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet	900
Db	3027	CATTCAAGGAACCGGAGCGCCACCACCGAGTGGAGCTGGCTTGGCCCCCTCTCTGAAGATG	3086
QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
Db	3087	GAAGGAACACCATTCATCGATACTCCAAGTTCTTGGAGCCCATCTGGAGACCCCAAA	3146
QY	921	ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu	940
Db	3147	CTCTCGGCCTTCTCAGCCTGTCCACACTTGTCTATGGGCCAAGCACAGCCCTTTGAATGAG	3206
QY	941	ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuSerIleAspLeuAspLysVal	960
Db	3207	ACTGCCCCCAAGTAACTCTTGGAAACATCAGAAGCTGCTCTCCATTGACCTGGCAAAAGTA	3266
QY	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer	980
Db	3267	GTCTTACCCCAACTTCGATCGAATCGCCCTCAAGTGAGACCTTTGTCCCTGGAGAAAGT	3326
QY	981	GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln	1000
Db	3327	GGTGCCTGGGACATTCCTGGAGGATCATCGCTGGCGGCTACAAACCAAGAGGTGGGCAG	3386
QY	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln	1020
Db	3387	ACCATCCCTGCTTTCCTTCCTCGGAGCATGGTGCCCTTGGCTCTTTTGTGGTACAG	3446
QY	1021	IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu	1040
Db	3447	ATCAGCAAGGCCAAAAGCCGCGAAGCGGAGGAGGCCAGGCAAGCGTCCACAGCTT	3506
QY	1041	MetGlnGlnValHisProProLysThrProSerVal	1052
Db	3507	ACACAGCAGACCCACCCCAAGGACCCCGCTCAGT	3542
RESULT 4			
AX024609			
LOCUS	3895 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0026348.		
ACCESSION	AX024609		
VERSION	AX024609.1	GI:10184745	
KEYWORDS	Rattus sp.		
SOURCE	Rattus sp.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1 (bases 1 to 3895)		
AUTHORS	Davignon,J., Chretien,M., Laaksonen,R., Seidah,N. and Marcinkiewicz,M.		
TITLE	Mammalian subtilisin/kexin isozyme ski-1: a proprotein		
JOURNAL	Patent: WO 0026348-A 1 11-MAY-2000; DAIGNON JEAN (CA) ; CHRETIEN MICHEL (CA) ; LAAKSONEN REIJO (CA) ; SEIDAH NABIL (CA) ; MARCINKIEWICZ MIECZYSLAW (CA) ; MONTREAL INST RECH CLINIQUES (CA)		
FEATURES	Location/Qualifiers		
source	1..3895		
	/organism="Rattus sp."		
	/db_xref="taxon:10118"		
CDS	418..3576		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC08993.1"		
	/db_xref="GI:10184746"		

TYGCVRGSGYKVCGRALSTSPASVPACANTLLSVTVOKRLNPNASVQIALIASA
RRLGPNVWPFQGGCKDLRLVQTLISYVQASLSPSTLTCPPYTHMPPYQVLTIG
GMPETLVNVTLLMGUWGLVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVW
SVYKASWEGIAQIHLMTVASPAREKLVNENUTSVLPLPQVLTTPSPRSKAVI
WQOYNLNPVPGVDPNNILKMDPLDWDJHVPNEDVAOHLRSKAFYFVWGLAPP
TCDATQVTLTLWDSERYPEETAKLREDYDNGLSLWFSQVWYNTSVWAKVKEVFE
NTQWMPDQDQGVNVALNELLVWNNWSSDGLYBGEELANHDYVAGSGZASRPE
DGVVLTQVTPDQGVLEKQVANDVNPVILGLVQVLPABGGGRVLVYDGSNGLDHSR
KDCVFLDALQVLTQVTPPSLSHSGNRQPPSGALAPERMEGNHLRYSVKLEA
HLGDKPRLPACPLSHWAKQFPOVNTAPSNLWKKHQLLSLDLDKVLPLNPSNRQV
RPLSPGSGANDIPGIMPRGYNEVQOTVPFATLGAMVALAFVVOIQISAKSRPKR
RRPRKBBQALQOAAHPAPSPSY*

Db 2218 CATACTTCACAGTGAAGTGCCTCAAGGTGAAGATCATTCCTCCACCCTCTCTCGAGC 2277
QY 621 LysArgValLeuTyrAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg 640
Db 2278 AGAGAGTCTCTGGGACCATGATACCAACCTCCGCTACCCACCCGGCTACTTCCCGAG 2337
QY 641 AspAsnLeuArgMetLysAsnAspProLeuAspTyrAsnGlyAspHisIleHisThrAsn 660
Db 2338 GACAACCTTCGGATGAAGATGATCTTTAGACTGGAATGGGACCATCTCCACACCAAC 2397
QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
Db 2398 TTCAGGGACATGTACCAGCATCTCGCAGCATGGCTACTTTGTGAGGTGCTTGTGTGCC 2457
QY 681 ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
Db 2458 CCATTCACATGCTTTGACCCGACGAGTACGGCACTCTGCTTAAGTGGACAGTAGGAA 2517
QY 701 GluTyrPheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu 720
Db 2518 GAGTACTTCCCTGAGGAGATGCTTAAGCTGAGGAGGAGCTGGACATGGCCTTCCCTT 2577
QY 721 ValIlePheSerAspTyrTrpAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
Db 2578 GTCGTCTTCAGTGACTGGTACACACTTCTGTATGAGAAAGTGAAGTTTACGATGAA 2637
QY 741 AsnThrArgGlnTrpThrMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGlu 760
Db 2638 ACACAAAGCAGTGGTGGATGCGAGATACATGAGGAGGACCAAGCTCCAGCTCTAAACGAG 2697
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
Db 2698 CTGCTCTGTGTGGAACATGGGTTCAGTGACGGCTGTATGAAGGGAGTTTGGCCGTG 2757
QY 781 AlaAsnHisAspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
Db 2758 GCACAAACAGCAGATGATGATGATGCGGTGTCAGCATTCGCCAGTTTCCAGAAAGTGT 2817
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
Db 2818 GTGGTATCACACAGACTTTCAGGACCAAGGATGGAAGTCTTAAACAAAGACAGACGA 2877
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
Db 2878 GTTGTGCAATGTCCTCCATCTTGGGGCTATATCAGATTCACATTCAGGTGGAGCCGG 2937
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
Db 2938 ATTGTGCTGTATGGAGACTCCAACTGCTTGGATGACAGTCACAGACAGAAAGGACTGCTT 2997
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
Db 2998 TGCCTTCTGGATGACTCTCTCAGTACATCTCATGGTGACCTCCAGCTCCAGCCCTCAGC 3057
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
Db 3058 CATTCAGGAACCGGCAGCGCCACCCAGCGGGCTGGCTGGCCCTCTCTGAAGGATG 3117
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
Db 3118 GAAGGAACACCTTCTCATGCTACTCCAAGTCTTGTAGGCCCCACCTTGGAGACCCGAAA 3177
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
Db 3178 CTCTGGCCCTTCCAGCCCTGTCACACTTGTCTGGGGCCCAAGCCACAGCTTTGAATGAG 3237
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
Db 3238 ACGGCACCCAGTAATCTTTGGAAACACCAAGAGCTCTCTCCATTGACCTGGACAAAGTA 3297
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980

Db 3298 GTGTTACCAACTTTCGCTCAAAATCGCCCTCAAGTGAAGACCTTTGCTCCCTGGAGAAAGT 3357
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
Db 3358 GGTGCTTGGGACATTCCTGGAGGATCATGCTTGGCGCTTACACCAAGGAGTAGGCCAG 3417
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
Db 3418 ACCATCCCTGTTTTCCTTCCCTTGGAGCATGCTGCGCTTCTTCCTGCTGATACAG 3477
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3478 ATCATGAAGCCCAAGACCGCGCGCAAGCGGAGGAGCCCAAGCGTCCACACTT 3537
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3538 GCACAGAGCCCAACCTGCAAGGACCCCGCTCAGT 3573

RESULT 5
AF094821 3895 bp mRNA linear ROD 25-APR-1999
LOCUS Rattus norvegicus subtilisin/kexin isozyme SKI-1 precursor, mRNA,
DEFINITION complete cds.
ACCESSION AF094821
VERSION AF094821.1 GI:4679094
KEYWORDS
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 3895)
AUTHORS Seidah,N.G., Mowla,S.J., Hamelin,J., Mamarbachi,A.M., Benjannet,S.,
Toures,B.B., Basak,A., Munzer,J.S., Marcinkiewicz,J., Zhong,M.,
Barate,J.C., Lazure,C., Murphy,R.A., Chretien,M. and
Marcinkiewicz,M.
Mammalian subtilisin/kexin isozyme SKI-1: A widely expressed
proprotein convertase with a unique cleavage specificity and
cellular localization
Proc. Natl. Acad. Sci. U.S.A. 96 (4), 1321-1326 (1999)
JOURNAL 99145548
MEDLINE 9990022
REFERENCE 2 (bases 1 to 3895)
AUTHORS Seidah,N.G., Mamarbachi,A.M., Hamelin,J. and Chretien,M.
Direct Submission
TITLE Submitted (24-SEP-1998) Biochemical Neuroendocrinology, Clinical
JOURNAL Research Institute of Montreal, 110 Pine Ave. West, Montreal, QC
H2W 1R7, Canada
FEATURES
source
1. .3895
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/cell_line="PC12"
/tissue_type="adrenal glands"
418. .3576
/note="subtilisin-like serine proteinase; cleaves after
hydroxylated amino acids such as Thr and possibly Ser"
/product="subtilisin/kexin isozyme SKI-1 precursor"
/protein_id="AAD27011.1"
/db_xref="GI:4679095"
/translation="MKLVNLIWLLLVLLCGKKHGLDRLGKKAFAKAPCPSCHLTLK
VEFSSTVVEYIVAEVGTAKARNSFISSALKSSEVDNWRILIPNNSDDYPSDFE
VQIKREKQAGLLTLEDHFNKRVTPQRVFRSLKFAESDPIVPCNETSRQKQSSR
PLKRASLSGSGFWHATGRHSRRLLRPRQAQTLQADVLQMGTGNSQKRWAVFD
TGLSEKHPFKNVKRTNERTLDDGLGHTFVAGVASMRECCGFAPDAELHIFR
VTNNQVSYTWNFLDAFNVAIKLKMVDLNSIGGPDFMDHPFDVKVWELTANNVIMVS
ATNGDPLGTLNPNADQMDVIGVGIGDFEDNIARFSRRMTTWELPGLPVKRPDIIV
TVGAVRGSGVGGCCALSGTSVAPVAGATLLVSTVQKRELNPASVQKALLASA
RRLPGVYMEFQGHKLLDRLAQLSYAPQASLSFYIDLECPITWMPYCSQPIYIG
GMPTIVNVTILMGMTGRIVDKPEWRYPQNGDNIEAFVSFYSVLWPMWPGYLAISI
SVTKKAASWEGIAQGHMITVASPAETELKNGAHTSTVKLPILKVKIITPTPRSKRLV

WDQYHNLRYPPGYPFRDLNRMKNDPLDWNQDGHVHTNFRMYOHLRSMGIFVEVLGAPF
 TCDDATQYGLTLLMVDSEEEYFPEETIAKLRDNDGLSLVVFSDWNTYNTSMRKVKFYDE
 NTRQWMPDTGANYPALNELLVWNMGFSDGLYEGEFALANDMYIYASGCSIAIRFPE
 DGVITQTFKQDGLVLEKQETAVNDVPTILGIYIPAEAGGRIVLYIGDSNGLDSDHRQ
 KDCFWLLDALQYTSYGVTPPSLSHSGNQRPPSGAGLAPPERMEGNHLHRYISVLEA
 HLDGPKRPLPACPHLSWAKPQPLNETAPSNLWKHOKLISIDLDKVLVLPNFRSRPOV
 RPLSPGESGAWDIPGIMPGRYNOEVGOTIPVFAFLGVALAFAFFVVOISRAKSRPKR
 RRPRAKRPQQAQAHAPARTPSV"

sig_peptide 418..468
 mat_peptide 469..3573
 misc_feature /product="subtilisin/kexin isozyme SKI-1"
 1069..1071
 /note="active site"
 1123..1125
 /note="glycosylation site"
 1162..1164
 /note="active site"
 1330..1332
 /note="glycosylation site"
 1429..1431
 /note="active site"
 1657..1659
 /note="active site"
 1957..1959
 /note="glycosylation site"
 2599..2601
 /note="glycosylation site"
 3232..3234
 /note="glycosylation site"
 3421..3480
 /note="transmembrane-region site"
 3481..3573
 /note="cytosolic tail; unclassified site"

BASE COUNT 952 a 991 c 1055 g 897 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3895
 Score: 5476.00 Matches: 1017
 Percent Similarity: 98.57% Conservative: 20
 Best Local Similarity: 96.67% Mismatches: 15
 Query Match: 97.49% Indels: 0
 DB: 10 Gaps: 0

US-09-830-837-6 (1-1052) x AF094821 (1-3895)

QY 1 MetLysLeuValAsnIleThrLeuLeuValValLeuLeuValLysLysHis 20
 |||||
 Db 418 ATGAAGCTCGTCACATCTGGCTTCTTCGCTGGTGGTTCCTGTGGGAAAGCAT 477
 |||||
 QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40
 |||||
 Db 478 CTGGGTGACAGCTGGGGAGAAAGCTTTTGAAGAGCCCATGCCCCAGCTGTCCAC 537
 |||||
 QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrIleValAlaPhe 60
 |||||
 Db 538 CTGACTTTGAAGTGGAAATCTCCTCAACTGTGTGGGAATATGAATATATGTGGCTTTC 597
 |||||
 QY 61 AsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSer 80
 |||||
 Db 598 AACGGATATTCACAGCAAGCTAGAACTCATTTATTTCAAGTGTCTATAAAGCAT 657
 |||||
 QY 81 GluValAspAsnTrpArgIleIleProArgAsnProSerSerAspTyrProSerAsp 100
 |||||
 Db 658 GAAGTGGACACTGGAGANTATACCTCGGAACAACCCATCTAGTACTACCTAGTATGAT 717
 |||||
 QY 101 PheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuThrLeuGluAspHis 120
 |||||
 Db 718 TTTGAGGTGATTCAGATMAAGAGACGAGGCGGGCTCTCACACTTGAGATCAC 777
 |||||
 QY 121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu 140
 |||||
 Db 778 CCAACATCAAGCGGGTGACACCCAGCGAAAGTCTTTTGGTTCCTCGAGTTTGTCTGAA 837

QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg 160
 |||||
 Db 838 TCCGACCCCATTTGTCCTCTGTAAGAGCCCGGTGGAGCAAGTGGCATGCACTACAGT 897
 |||||
 QY 161 ProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpPheIleAlaThrGlyArgHis 180
 |||||
 Db 898 CCCCCTGAAAAGAGCCAGTCTCTCCCTGGGCTCTGGATCTTGGCATGCAACAGGAGCAT 957
 |||||
 QY 181 SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
 |||||
 Db 958 TCAAGTCGACGATTTGCTGAGAGCCATCTCTCGCCAGGTTGCCACAGCATTCAGCGCAGAT 1017
 |||||
 QY 201 ValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValAlaPheAspThrGly 220
 |||||
 Db 1018 GTGCTTTGGCAGATGGGATACACAGGTCTAATCTCAGGGTTGCCGCTTTTGCATCTGGG 1077
 |||||
 QY 221 LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240
 |||||
 Db 1078 CTCAGTGAGAGCATCCACATTTCAAGAATGTGAAGAAAGAACCACTGGACCAATGAG 1137
 |||||
 QY 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
 |||||
 Db 1138 CGGACCTGGACCATGGCTGGCCATGCCACATCTGTCAGGTGTGATTCGCGCATG 1197
 |||||
 QY 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
 |||||
 Db 1198 AGAGATGCCAAGATTTGCCCATGCCAGATGCCACATCTCAGGGTCTTTACCAAC 1257
 |||||
 QY 281 AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLys 300
 |||||
 Db 1258 AATCAGGTGCTTACAGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1317
 |||||
 QY 301 IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisProPheValAsp 320
 |||||
 Db 1318 ATGGAGCTTCAACCTTACATCGGTGGCTGACTCATGATGATGATGATGATGATGATG 1377
 |||||
 QY 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
 |||||
 Db 1378 AAGGTATGGGAATTAACAGCGAACAAATTAATCAATGTTCTCTATTGGCAATGATGGA 1437
 |||||
 QY 341 ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 360
 |||||
 Db 1438 CCTCTCTATGGCACTCTGAATAACCTCTGATCAGATGATGATGATGATGATGATGATG 1497
 |||||
 QY 361 IleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrThrTrpGluLeu 380
 |||||
 Db 1498 ATTGACTTTGAAGACAAACATCGCCGCTTCTCTCCAGGGGAATGACTACCTGGGAATA 1557
 |||||
 QY 381 ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly 400
 |||||
 Db 1558 CCGGAGCGCTATGCTGCTGAGGCTGACATTTGCACCTATGCTGCTGAGTGGGGGT 1617
 |||||
 QY 401 SerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420
 |||||
 Db 1618 TCTGGTGTGAAGGGGCTGCCGTCGCTCAGGACCTCTCAGGGACAGTGTGCCCTCCCGAGTGT 1677
 |||||
 QY 421 AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgIleValValAsnProAla 440
 |||||
 Db 1678 GCTGGGGCTGTACCTTTGTAGTAAAGCAGTACAGAGCGGGAGGTAGTGAATCCTGCC 1737
 |||||
 QY 441 SerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAlaAsnMetPheGlu 460
 |||||
 Db 1738 AGTGTGAAGCAAGCTTTGTAGTATGATCAGCCCGGAGACTTCTGCTGTCAACATGTTTGA 1797
 |||||
 QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480
 |||||
 Db 1798 CAAGCCATGCCAAGTTGGATCTACTTGGGAGCCTATCAGATCTCTCAGCAGCTATAACCG 1857
 |||||
 QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
 |||||
 Db 1858 CAGGCGAGCCTGAGTCTCTAGCTACATCCAGCTGACTGAGTGTCCCTACATGTGGCCCTAC 1917

QY	501	CysSerGlnProIleTyrTyrGlyGlyMetProThrValValAsnValThrIleLeuAsn	520
Db	1918	TGCTCCAGCCCACTACTATGAGGAATGCCAACAAATGTTAATGTCACCAATCCTCAAT	1977
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln	540
Db	1978	GGCATGGGAGTTACAGAGAAGATTGGATAAGCCTGAGTGGGACCCCTATTATTACCACAG	2037
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTrpSerGly	560
Db	2038	AATGGAGACAACATTCAAGTGGCCTTCCTACTCCTCAGTGTGGCCTTGTCAGGT	2097
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln	580
Db	2098	TACCTTGGCCATCTCCATTTCTGTGACAAAGAGCCAGCTTCTCGGAAGGCATCGGCAG	2157
QY	581	GlyHisValMetIleThrValAlaLaserProAlaGluThrGluSerLysAsnGlyAlaGlu	600
Db	2158	GGCCACATCATGATCAGTGGCTTCCCAGCAGAGACCGAAATTAATAAATGGTGGCGAG	2217
QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProArgSer	620
Db	2218	CATACTTCCACAGTGAAGTGGCCATCAAGTGAAGTCAATTCACACCCCTCTCTCGAGC	2277
QY	621	LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg	640
Db	2278	AAAGAGAGTCTCTGGGACCAAGTACCACAACCTCCGCTACCCACCCGGCTACTTCCCGAG	2337
QY	641	AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660
Db	2338	GACAACTTCGGGATGAAGTAATGATCCTTTAGACTGGAATGGGACCACTCCACACCAAC	2397
QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
Db	2398	TTACAGGACATGTACCAGCATCTGCCAGCATGGGCTACTTGTGGAGGTGCTTGGTGCC	2457
QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
Db	2458	CCATTTCACATGCTTTGACGCCACGAGTACGGCACTCTGTTATGGTGACAGTAGGAA	2517
QY	701	GluTyrPheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu	720
Db	2518	GAGTACTTCCCTGAGGAGATGCTTAAGCTGAGGAGGACGTGACAAATGGCCTTCCCTT	2577
QY	721	ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
Db	2578	GTGCTCTTCAGTGACTGGTACAACTTCTGTTATGAGAAAAGTGAAGTTTACGATGAA	2637
QY	741	AsnThrArgGlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGlu	760
Db	2638	AAACAAGGACGTGGTGGATGCCAGATACCTGGAGGAGCAACGCTCCAGCTCTAAACGAG	2697
QY	761	LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu	780
Db	2698	CTGCTGCTGTGTGGAACTAGGGTTTCAGTGAGGGCTGTATGAAGGGAGTTTGGCCGTG	2757
QY	781	AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800
Db	2758	GCAAAACCGACATGATCATGCATCGGGTCAGCATTCGCCAGTTTCCAGAGATGAT	2817
QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla	820
Db	2818	GTGGTGATCACACAGACTTTCAGGACCAAGGATTTGGAAGTCTTAAACAAGACAGCA	2877
QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
Db	2878	GTGTGCGAACAATGTCCTCCCATTTCTGGGGCTATATCATGATTCAGATTCAGCTGAAGGTGGAGCCGG	2937
QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
Db	2938	ATTGTGCTGTATGGAGACTCCCACTGCTTGGATGACAGTACACAGACAGAGGACTGCTT	2997
QY	861	TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer	880

Db	2998	TGCTTCTTGATGACCTCTTTCAGTACACATCCTATGGTGTGACCCCTCCAGCCTCAGC	3057
QY	881	HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet	900
Db	3058	CATTAGGGAACCGGACGCCACCCAGCGGGCTGGCTTGGCCCCCTCTCTGAAGATG	3117
QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
Db	3118	GAAGGAAACCACTTCATCGCTACTCCAAAGTTCTTGAGGCCACCTTGGGAGACCCGAA	3177
QY	921	ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu	940
Db	3178	CCTCGGCCCTTCCAGCCTCTGCACACATTCCTGGGCCAAGCCACAGCCTTTGAATGAG	3237
QY	941	ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuSerIleAspLeuAspLysVal	960
Db	3238	ACGGCACCCAGTAACTTTTGGAAACACCAAGCTGCTCTCCATTTGACCTGGACAAAGTA	3297
QY	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer	980
Db	3298	GTGTTACCCAACTTTCGCTCAATCGCCTCAAGTGAGACCTTGTCCCTGGAGAAAT	3357
QY	981	GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln	1000
Db	3358	GGTGCCTGGACATTCCTGGAGGATCATGCTCGCGCTTACAAACAGGAAAGTAGGCCAG	3417
QY	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln	1020
Db	3418	ACATCCCTGTTTTGCTTTCCTGGAGCCTTGGTGGCCCTGGCTTCTTCGGGTACAG	3477
QY	1021	IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu	1040
Db	3478	ATCAGTAGGCCAAGACCGCGCGAAGCGGAGGAGGCCAGGCGCAAGCGTCCACAACTT	3537
QY	1041	MetGlnGlnValHisProProLysThrProSerVal	1052
Db	3538	GCACAGCAGCGCCACCTCGCAGGAGCCCGCTCAGTG	3573

RESULT	6
LOCUS	AX024611
DEFINITION	Sequence 3 from Patent WO0026349.
ACCESSION	AX024611
VERSION	AX024611.1
KEYWORDS	GI:10184747
SOURCE	Mus sp.
ORGANISM	Mus sp.
REFERENCE	AX024611
AUTHORS	Sequence 3 from Patent WO0026349.
TITLE	Sequence 3 from Patent WO0026349.
JOURNAL	Patent: WO 0026349-A 3 11-MAY-2000;
FEATURES	DAVIGNON JEAN (CA) ; CHRETIEN MICHEL (CA) ; LAAKSONEN REIJO (CA) ;
source	SEIDAH NABIL (CA) ; MARCINKIEWICZ MIECZYSLAW (CA) ; MONTREAL INST
CDS	RECH CLINIQUES (CA)
	Location/Qualifiers
	1..3788
	/organism="Mus sp."
	/db_xref="taxon:10095"
	8..3166
	/note="unnamed protein product"
	/codon_start=1
	/protein_id="CAC08994.1"
	/translation="MKLVSTWLVLLVLLCGKRLGDLRTRALEKAPCPSCHLTLK
	VFSSNVVEVETVAENGFTAKARNSFSSALKSSEVENWRIIPNNSSDYPSPDFE
	VIOKEKQAGLLTLEDHPNKRTPORVFRSLKFAESNPVPCNTRWSQKWSR
	PLIKRSGSGFWHTAEDRHSSRLRLRAIPRQAQTLQADVLQMGTGTGANRVAVFD
	TGLSEKHPFNKYNRTNERTLDGGLGHGTFVAGVTSASRECOGFPADLEHIFR
	VFTNNQSVYTSWFLDAFNAILKMDVLNLSIGGPDFMDHPFVDKWEIETANNVIMS

Qy	601	GlnThrSerThrValLysLeuProIleLysValLysIleLeuProThrProArgSer	620	Db	2888	GTGTTTACCCAACTTTCGATCCAATCGCCCTCAAGTGAGAGACTTGTCCCTCGGAGAGT	2947
Db	1808	CACACTTCCACCGTGAAGCTGCCATCAAGGTGAAGATCATTCACACCCCTCTCCGAGC	1867	Qy	981	GlyAlaTrpAspIleProGlyGlyMetProGlyArgTyraGlnGluValGlyGln	1000
Qy	621	LysArgValLeuTrpAspGlnIleHisAsnLeuArgTyraProGlyTyraPheProArg	640	Db	2948	GGTGGCTGGGACATTCCTGGAGGGATCATGCTGGCCGCTACACACGAGAGGTGGACAG	3007
Db	1868	AAGAGAGTCTCTGGGACCACTACACACCTCGCTACCCACCTGGCTACTTCCCCAGG	1927	Qy	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGln	1020
Qy	641	AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660	Db	3008	ACCATCCCGCTTTCGCCCTTCTCGGAGCCATGTGGCCCTTGTGTGGTACAG	3067
Db	1928	GACAACTTCGGGATGAAGATGACCTTTAGACTGGATGGCGACACGCTCCACCAAC	1987	Qy	1021	IleAsnLysAlaLysSerArgProLysArgLysProArgValLysArgProGlnLeu	1040
Qy	661	PheArgAspMetTyrcGlnHisLeuArgSerMetGlyTyraPheValGluValLeuGlyAla	680	Db	3068	ATCAGCAAGGCGCAAGCGCCGAGGAGGAGGCCAGGCGCAAGCGTCCACAACTT	3127
Db	1988	TTTCAGGAGCATGTACAGCATCTGGCAGCATGGCTACTTCTGGAGGCTCTCGGCGCC	2047	Qy	1041	MetGlnGlnValHisProProLysThrProSerVal	1052
Qy	681	ProPheThrCysPheAspAlaSerGlnIleTyraGlyThrLeuLeuMetValAspSerGluGlu	700	Db	3128	GCACAGCAGGCCACCCCTGCAAGAGCCCATCATCATG	3163
Db	2048	CCATTACATGTTTACGCCACACAGATGTCATCTTGCTGCTGGTGAGAGTGAGGAA	2107	RESULT 7			
Qy	701	GluTyraPheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu	720	LOCUS	AF094820	3788 bp mRNA linear	25-APR-1999
Db	2108	GAGTACTTCCCTGAGGAGATTCTAAGCTGAGGAGGATGTGGACATGGCCCTTCCCTC	2167	DEFINITION	Mus musculus subtilisin/kexin isozyme SKI-1 precursor, mRNA,		
Qy	721	ValIlePheSerAspTrpTrpAsnThrSerValMetArgLysValLysPheTyraAspGlu	740	ACCESSION	complete cds.		
Db	2168	GTCACTTTCAGTCACTGGTACAACTCTCTGTATGAGAAAAGTGAATTTATGATGAA	2227	VERSION	AF094820		
Qy	741	AsnThrArgGlnTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGlu	760	KEYWORDS	AF094820.1 GI:4679092		
Db	2228	ACACACGAGCATGTGGATGCCACACACCGGAGGAGGAGCAATCCACGCTCTGAATGAG	2287	SOURCE	Mus musculus.		
Qy	761	LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyraGluGlyGluPheThrLeu	780	ORGANISM	Mus musculus.		
Db	2288	CTGCTCTGTGTGGAACATGGGTCTAGTGACGGCCATATGAGGGGAGTGTCTCTG	2347	REFERENCE	1 (bases 1 to 3788)		
Qy	781	AlaAsnHisAspMetTyraAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800	AUTHORS	Seidah,N.G., Mowla,S.J., Hamelin,J., Mamarbachi,A.M., Benjannet,S., Toureaux,B.B., Basak,A., Munzer,J.S., Marcinkiewicz,J., Zhong,M., Barale,J.C., Lazure,C., Murphy,R.A., Chretien,M., and Marcinkiewicz,M.		
Db	2348	GCAAAACATGACATGACTATGCGTCGGGTGCAGCATCGCAAGTTCACAGAGATGCG	2407	TITLE	Mammalian subtilisin/kexin isozyme SKI-1: A widely expressed proprotein convertase with a unique cleavage specificity and cellular localization		
Qy	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnIleThrAla	820	JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (4), 1321-1326 (1999)		
Db	2408	GTGCTGATCACACAGCTTCAAGGACCAAGATGGAGTCTTAAACAGAGACAGCA	2467	MEDLINE	99145548		
Qy	821	ValValGluAsnValProIleLeuGlyLeuTyraGlnIleProAlaGluGlyGlyArg	840	PUBMED	9990022		
Db	2468	GTGTGGAAATGTTCCCATTTTGGGCTTTATCAGATTCATCTGAGGTGGAGCGCG	2527	REFERENCE	2 (bases 1 to 3788)		
Qy	841	IleValLeuTyraGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860	AUTHORS	Seidah,N.G., Mamarbachi,A.M., Hamelin,J. and Chretien,M.		
Db	2528	ATCGTCTGTATGGAGACTCCAACTGCTGGATGACAGTCACAGACAGAGGAGTCTTT	2587	TITLE	Submitted (24-SEP-1998) Biochemical Neuroendocrinology, Clinical Research Institute of Montreal, 110 Pine Ave. West, Montreal, QC H2W 1R7, Canada		
Qy	861	TrpLeuLeuAspAlaLeuLeuGlnIleTyraThrSerTyraGlyValThrProProSerLeuSer	880	FEATURES	Location/Qualifiers		
Db	2588	TGCTTCTGGATGCGCTCTCAGTACACATCTTATGGCGGACCCCTCCAGCCTCAGC	2647	source	1. 3788 /organism="Mus musculus" /db_xref="taxon:10090" /cell_lines="ACT20"		
Qy	881	HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet	900	CDS	8. 3166 /note="subtilisin-like serine proteinase; cleaves after hydroxylated amino acids such as Thr and possibly Ser" /codon_start=1 /product="subtilisin/kexin isozyme SKI-1 precursor" /protein_id="AAD27010.1"		
Db	2648	CATTACAGGAACCGGACGCCACTAGCGAGCGCGCTGGCCCTCTCTGAAGAGATG	2707		/db_xref="GI:4679093" /translation="MKLVSTWLLVVLGCKRHLGDRGTRALEKAPCPGCSHTLTK /translating="MKLVSTWLLVVLGCKRHLGDRGTRALEKAPCPGCSHTLTK VFQSTVVEYIVAFNGYTKARNSFSSALKSEVENWRIIPRNSSQPSDF VIOIKQKAGLLTLEDPNTKRTVPRKVFSLKFAENPIVPCNTRWSQWSSR TYGARGSGVKGCALSGTSVASPVAGAVTLLYSTQKRELVPNPASVKOALASA RRLPCVNMFOGCHGLDLRAYOILSSYKQASISPSYIDLTECPYWPQSYLYG SPTIVVNTILNGMGVTGRIVDPWRPYPQLPQGDNIIEAFSYSSVILPWSGTALSI CMYTKAASWEGIAQGLIMITVASPAETELSHGAETSTVAKPLIKVLIIPPRSKRLV WDQHNLRPPGYPFRDNLKMDPLDNLWDGHDVHTFRDMYQHLRSMGFVEVLGAP TCFATQGTLLVDSSEYFPFEEIAKLRRDNDGLSLVTSDFWNTSYMRKVKFDE TCFATQGTLLVDSSEYFPFEEIAKLRRDNDGLSLVTSDFWNTSYMRKVKFDE		
Qy	901	GluGlyAsnHisLeuHisArgTyraSerLysValLeuGluAlaHisLeuGlyAspProLys	920				
Db	2708	GAAAGAAACACCTCACTCGGTACTCCAAAGTCTTGAAGCCCACTTGGGAGACCCGAA	2767				
Qy	921	ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu	940				
Db	2768	CTCTGGCCCTGCCAGCTGTCCCATTTGTATGGCCCAAGCCACACGCTCTTGAATGAG	2827				
Qy	941	ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal	960				
Db	2828	ACGGACCCAGTAATCTTTGGAAACATCAGAACTGCTCTCAATGACCTGGACAAAGTA	2887				
Qy	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyValSer	980				

NRQWMPDGTGANIPALNELLVNMGMFSDGLYEGEFVLNHDYIAGSGSIARFPE
 DGVITQTFKDOGLEVKQETAVVENPILGLYQIPSEGGRIYLVGSDNCLDSDHQ
 KDFWLLDALQYTSYGVTPPSLSHSGNRQRPSPGAGLAPPERMEGNHLHRYSKVLEA
 HLDGPKRPLPACPHLSWAKPQPLNETAPSNLWKHOKLLSIDLDKVLNPNFRSNRPQV
 RPLSPGESGAWDIPGGIMPGRYNOEVQGTIPVFAFLGAMVALAFVYVQISAKSRPKR
 RRPRAKRPQLAQAHPARTPSV"

sig_peptide
 mat_peptide

misc_feature

misc_feature
 /note="experimentally shown not to be glycosylated in human and rat SKI-1; glycosylation site"

misc_feature

misc_feature
 /note="active site"

misc_feature

misc_feature
 /note="glycosylation site"

misc_feature

misc_feature
 /note="active site"

misc_feature

misc_feature
 /note="active site"

misc_feature

misc_feature
 /note="glycosylation site"

misc_feature

misc_feature
 /note="glycosylation site"

misc_feature

misc_feature
 /note="transmembrane-region site"

misc_feature

misc_feature
 /note="cytosolic tail: unclassified site"

BASE COUNT

915 a 967 c 1015 g 891 t

ORIGIN

Alignment Scores:
 Pred No.: 0 Length: 3788

Score: 5427.00 Matches: 1008

Percent Similarity: 98.19% Conservative: 25

Best local Similarity: 93.82% Mismatches: 19

Query Match: 96.62% Indels: 0

DB: 10 Gaps: 0

US-09-830-837-6 (1-1052) x AF094820 (1-3788)

QY 1 MetLysLeuValAsnIleTrpLeuLeuValValLeuLeuCysGlyLysHis 20

DB 8 ATCAAGCTCTCAGCACCTGGCTTCTGCTGGTGGTTCCTCTGTGGAAAGCGCAC 67

QY 21 LeuGlyAspArgLeuGlyLysSerPheGlyLysAlaProCysProGlyCysSerHis 40

DB 68 CTGGCGACAGGCTGGGACGAGAGCTTGGAAAGCGCGCTGCCACGCTGCCAC 127

QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrIleValAlaPhe 60

DB 128 CTGACTTTGAAGTGGAAATCTCTCAACTGTGGTGGAGTACGAATATATTGGCTTTC 187

QY 61 AsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSer 80

DB 188 AACGGATCTTACACCCAAAGCTAGAACTCATTTATTTCAAGTGGCTGGAAAGCACT 247

QY 81 GluValAspAsnTrpArgIleProArgAsnAsnProSerSerAspTyrProSerAsp 100

DB 248 GAAGTGGAAACTGGAGATATACCTCGGAACACCCATCCAGTACTACCTAGTATGAT 307

QY 101 PheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis 120

DB 308 TTTGAGTGTATCAGATATAAAGAGAGAGAGCGGGGCTGCTCACACTTGAAGATCAC 367

QY 121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu 140

DB 368 CCCAACATCAAGCGGGTGACACCCAGCGGAAAGCTCTTTGTTCCCTCAAGTTGCTGAA 427

QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg 160

DB 428 TCCAAACCCATCGTGCCTGTATGAAACCCGGTGGAGCCAGAAAGTGGCAGTCACTACGT 487

QY 161 ProLeuArgArgAlaSerLeuGlySerGlyPheTrpPheHisAlaThrGlyArgHis 180

DB 488 CCCCTGAAAGAGCCAGTCTCTCCCTGGCTCTGGATCTGGCATGCAACAGGAAGACAT 547

QY 181 SerSerArgGluLeuArgAlaIleProArgGlnValAlaGluThrLeuGlnAlaAsp 200

DB 548 TCNAGTGGCGATTTGTAGAGCCATCTCTCCCGGGTCCGCCACACTGCAGGCAGAT 607

QY 201 ValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGly 220

DB 608 GTGCTGTGGCAGATGGGATACACAGGTCTAATGTACAGTTGCTGTTTGTACTGCGG 667

QY 221 LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240

DB 668 CTCAGTGAGAAGCATCCGCATTTTAAGAAATGTGAAGAGAGAACCAACTGCACCAATGAG 727

QY 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260

DB 728 CGGACCTGGATGATGGCTAGGCCATGGCACATTCGTTCCAGGTGTGATTCGCAGCATG 787

QY 261 ArgGlyCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280

DB 788 AGGGAGTGCCAAAGATTTGCTCCAGATGCAGAGCTGCACATCTTCAGGGTCTTTACCAAC 847

QY 281 AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLys 300

DB 848 ATCAGGTGTCTTACACATCTTGGTTCTGGATGCTTCAACTATGATCCATCCTAAAGAAG 907

QY 301 IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisProPheValAsp 320

DB 908 ATGGAGCTTCTCAACCTTAGCATCGTGGCGCCGACTTCATGGATCATCCGTTGTGTAC 967

QY 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340

DB 968 AAGGTGGGAAATTAACAGTAACTAATATATATGTTTCTGCTATTTGGCAATGATGGA 1027

QY 341 ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 360

DB 1028 CCTCTCTATGGCACTCTGAATTAACCTGCTGATCAGATGATGATTTGGAGTGGGTGGC 1087

QY 361 IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGluLeu 380

DB 1088 ATTGACTTTGAAGATAACATCGCTCGCTTTTCTCCAGGGGAATGACTACCTGGGAATTA 1147

QY 381 ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly 400

DB 1148 CCAGGAGCTATGCTGCTGTAAGCTGCACATTCACCTATGCTGCTGGAGTGGGGGT 1207

QY 401 SerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420

DB 1208 TCCGGTGTGAAGGGGCTGCGTGCCTCTCAGGACCACTGCTCGCTTCCCAAGTGGTC 1267

QY 421 AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAla 440

DB 1268 GCTGGGCGCTACCTTTGTAGTAAAGACAGTACAGAGCGGGAGCTGGTGAATCTGTC 1327

QY 441 SerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGlu 460

DB 1328 AGTGTGAAGCAAGCTTTGATGACGCCCGAGACTCTCTGGGGTCACTATGTTGAG 1387

QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480

DB 1388 CAAGGTATGCAAGTGGATGCTGTCGAGCTTATCAGATCTCAGCAGGTATAAACCG 1447

QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500

DB 1448 CAGGCAAGCTGAGTCTAGCTACATCGACCTGACTGAGTGTCTCTACATGTGCGCCTAC 1507

QY	501	CysSerGlnProIleTyrGlyMetProThrValValAsnValThrIleLeuAsn	520
Db	1508	TGCTCCAGCCCTATCTACTATGAGGAATGCCAACAATCGTTAATGTCACCATCCTCAAT	1567
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrIleuProGln	540
Db	1568	GGCATGGGGGTGCACAGGAAGAATTGTGATAAGCCTGAGTGGCGACCCCTATTATCCACAG	1627
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerValIleuTrpProTrpSerGly	560
Db	1628	AATGGAGACAACTGAAGTGCCCTCTCTACTCTCAGTGTGTGGCCCTGGTCAGGT	1687
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln	580
Db	1688	TACCTTGCCATCTCCATTTCTGTGACCAAGAAGGCAGCTCTCTGGGAAGGCATCGCTCAG	1747
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600
Db	1748	GGCCCACTCATGATCACAGTGGCGTCCCCAGAGACAGAGATTACACAGTGTGGCGAG	1807
QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProArgSer	620
Db	1808	CACACTTCACCGTGNAGCTGCCATCAAGGTGAAGATCATCTCCACCCCTCTCGGAGC	1867
QY	621	LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg	640
Db	1868	AAGAGAGTCTCTGGGACCCAGTACCACAACTCCGCTACCCACCTGGCTACTTCCCGAGG	1927
QY	641	AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660
Db	1928	GACAACTTCGGATGAAGAATACCCCTTTAGACTGGAAATGGCGACCATCTCCACACCAAC	1987
QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
Db	1988	FTCAGGGACATGTACCAGCATCTGGCCACCATGGGCTACTTCTGTGGAGGTGCTGGCGCC	2047
QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
Db	2048	CCATTTCACATGTTTGGACCCACACAGTATGGCACTTCTGCTGTGTGCACAGTGAAGAA	2107
QY	701	GluTyrPheProGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu	720
Db	2108	GAGTACTTCCCTGGAGAGATTCTAAGCTGAGGAGGGATGTGGACAATGGCCCTTCCCTC	2167
QY	721	ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
Db	2168	GTCATCTTCAGTACTGGTACAACTTCTGTATGAGAAAGTGAAGTTTATGATGAA	2227
QY	741	AsnThrArgGlnTrpTrpMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu	760
Db	2228	AACACAGGACGTGTGGATGCCAGACACCGGAGGAGCGAATCCACGCTCTGAATGAG	2287
QY	761	LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu	780
Db	2288	CTGCTGTCTGTGGAACATGGGGTTCACTGACGGCTATATGAAGGGAGTTTGTCTGT	2347
QY	781	AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800
Db	2348	GCAAACCATGACATGATCTATGCTGGGGTGCAGCATGCCAAGTTTCCAGAAGATGCC	2407
QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla	820
Db	2408	GTCTGTCATCACACAGCTTTCAGGACCAAGATGGAGTCTTAAACAAGACAGACAGCA	2467
QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
Db	2468	GTGTGGAAATGTTCCCATTTTGGGGCTTTATCAGATTCCATCTGAAGTGGAGGCCGG	2527
QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
Db	2528	ATCGTGCTGTATGGAGACTCCAACTGCTTGGATGATGACAGTACAGACAGAAAGACTGCTTT	2587
QY	861	TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer	880
Db	2588	TGGCTTCTGGATGCGCTCTTTCAGTACACATCTATGCGTGCACCCCTCCAGCCTCAGC	2647
QY	881	HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet	900
Db	2648	CATTGAGGAACGGCGAGCCACCTAGCGGAGCGGCTTGGCCCTCTCTGAAGAGATG	2707
QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
Db	2708	GAAGGAACACCTCCATCGGTACTCCAAAGTTCTTGAAGCCACCTTGGAGACCCGAAA	2767
QY	921	ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu	940
Db	2768	CCTGGGCCCTGCCAGCCTGTCCACATTTGTTCATGGGCCCAAGCCACAGCCTTTGAATGAG	2827
QY	941	ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal	960
Db	2828	ACGGCACCCAGTAATCTTTGGAACATCAAGAGCTCTCTCCATTGACCTGGACAAGTA	2887
QY	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer	980
Db	2888	GTGTATACCACTTTCATCCCAATCCCTCAAGTGAGAGACCTTTGTCCCTGGAGAGAGT	2947
QY	981	GlyAlaTrpAspIleProGlyLysIleMetProGlyArgTyrAsnGlnValGlyGln	1000
Db	2948	GGTGGCTGGACATCTCTGGAGGGATCATGCTGGCGCTACAAACAGGAGGTGGGACAG	3007
QY	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln	1020
Db	3008	ACCATCCCTCTTCTCGCTTCTCGGAGCATGTTGGCCCTGGCCCTTCTTGTGGTACAG	3067
QY	1021	IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu	1040
Db	3068	ATCAGCAAGCCCAAGAGCGCGCGAAGCGAGGAGGCCCAAGGCGTCCACAACTT	3127
QY	1041	MetGlnGlnValHisProProLysThrProSerVal	1052
Db	3128	GCACAGCAGCGCCACCCCTGCAAGGACCCCATCATGTG	3163
RESULT	8		
LOCUS	BC011533	4204 bp mRNA linear	ROD 07-AUG-2002
DEFINITION	Mus musculus, Similar to membrane-bound transcription factor protease, site 1, clone MGC:6457 IMAGE:2615736, mRNA, complete cds.		
ACCESSION	BC011533		
VERSION	BC011533.1	GI:15079376	
KEYWORDS	MGC.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 4204)		
AUTHORS	Straussberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgabbs-r@mail.nih.gov Tissue procurement: Jeffrey Green M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMML) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amg@bcm.tmc.edu Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.		


```
Db 1786 CAAGGTGATGGCAAGTTGGAT-----CTGGCAGCTTATCAGATCCTCAGCAGCTATAAACCG 1842
QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
Db 1843 CAGGCAAGCCTGAGTCTAGCTACATCGACCTGACTGAGTGTCCCTTACATGTGGCCCTAC 1902
QY 501 CysSerGlnProIleTyrTyrglyMetProThrValValAsnValThrIleLeuAsn 520
Db 1903 TGTCCAGCCTATCTACTATGAGGAATGCCAACATCGTTAATGTCAACCATCCTCAAT 1962
QY 521 GlyMetGlyValThrGlyArgIleValAspLysProAspTrpClnProTyrLeuProGln 540
Db 1963 GGCATGGGCGTACAGAGAAATGTGGATAAGCCTGAGTGGCGACCCCTATTATACCACAG 2022
QY 541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTrpSerGly 560
Db 2023 AATGGAGACAACATTCAAGTGGCTTCTCTACTCTCAGTGTGTGGCCCTGGTCAGT 2082
QY 561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluIleAlaGln 580
Db 2083 TACCTTGGCATCTCCATTCTGTGACCAAGAGGCAGCTTCTGGGAAGGCATCGCTCAG 2142
QY 581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu 600
Db 2143 GGCACATCATGATCAGTGGCGTCCCGCAGACAGACAGATTACACAGTGGTGGCGAG 2202
QY 601 GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProArgSer 620
Db 2203 CACACTTCCACCGTGAAGTGGCCCATCAGGTGAAGATCATTCCTCCACCCCTCTCTGGAGC 2262
QY 621 LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg 640
Db 2263 AAGAGAGTCTCTGGGACCAGTACCACAACCTCCGCTACCCACTGGCTACTTCCCGAG 2322
QY 641 AspAsnLeuArgMetLysAsnProLeuAspTrpAsnGlyAspHisIleHisThrAsn 660
Db 2323 GACAACCTCGGGATGAAGATGACCTTTAGACTGGGAATGGGACACACCTCCACCAAC 2382
QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
Db 2383 TTCAGGGACATGTACCGAGATCTGCGCAGATGGGCTACTTCTGGAGGTGCTCGGCGCC 2442
QY 681 ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
Db 2443 CCATTTCACATGTTTGACGCCACACAGTATGGCACTTTCGCTGCTGGTGACAGTGAGGAA 2502
QY 701 GluTyrPheProGluLeuAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu 720
Db 2503 GAGTACTTCCCTGAGGAGATGCTAAGCTGAGGAGGATGTGGACAATGGCCCTTCCCTC 2562
QY 721 ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
Db 2563 GTCATCTTCAGTGACTGGTACACACTTCTGTATGAGAAAGTAGTATATGATGAA 2622
QY 741 AsnThrArgGlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGlu 760
Db 2623 AACACAGGACAGTGTGTGATGTCAGACACCGGAGGAGCAACATCCAGCTCTGAATGAG 2682
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyClnPheThrLeu 780
Db 2683 CTGCTGTCTGTGTGAACATGGGTTCAGTGAAGCCCTATATGAAGGGGAGTTTCTCTG 2742
QY 781 AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
Db 2743 GCAACCATGACATGACTACTATGCTGGGGTGCACATCGCCCAAGTTTCCAGAAATGCG 2802
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
Db 2803 GTCGTGATCACACAGACTTTCAGGACCAAGGATTTGGAGGTCTTAAACACAGACAGCA 2862
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
```

```
2863 GTTGTGGAAAATGTTTCCCATTTTGGGGCTTTATCAGATTCCATCTGAAGTGGAGCCGG 2922
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysaspCysPhe 860
Db 2923 ATCGTCTGTATGGAGACTCCAACCTCTTGGATGACAGTACACAGACAGAGGACTGCTT 2982
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
Db 2983 TGGCTTCTGATGCGCTCTCTCAGTACACATCCTATGGGTGACCCCTCCAGCCTCAGC 3042
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
Db 3043 CATTCAGGAACCGGCGAGCCACCTAGCGAGCGGCTTGGCCCTCTCTCAAAGATG 3102
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
Db 3103 GAAGGAACACCTCCATCGGTACTCCAAGTCTTGAAGCCACCTTGGGAGACCCGAA 3162
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
Db 3163 CCTCGGCCCTGCGAGCTGTCCACATTTGTATGGGCCAAGCCACAGCCTTTGAATGAG 3222
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLysVal 960
Db 3223 ACGGCACCCAGTAATCTTTGGAACATCAGAAGCTGCTCTCCATTGACCTGACAAAGTA 3282
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3283 GTGTATACCAACTTTCGATCCCAATCGCCTCAAGTGAGACCTTTGTCCTCGGAGAGT 3342
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
Db 3343 GGTGCTGGGACATCTCTGGAGGATCATGCTGGCGCTACAAACCAGGAGTGGGACAG 3402
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGln 1020
Db 3403 ACCATCCCGCTTCTGCTTCTCGGAGCATGTTGGCCCTTCTTGTGTGTGATACAG 3462
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3463 ATCAGCAGCCGACAGAGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3522
QY 1041 MetGlnGlnValHisProLysThrProSerVal 1052
Db 3523 GCACAGCAGGCCACCCCTCGAAGGACCCCATCATG 3558

RESULT 9
BC026330
LOCUS
DEFINITION
Homo sapiens, similar to membrane-bound transcription factor
protease, site 1, clone MGC:26368 IMAGE:4816102, mRNA, complete
cds.
ACCESSION BC026330
VERSION BC026330.1 GI:20072867
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3026)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```


DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 32 Row: m Column: 11
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4506774.

FEATURES

source

1. 3026

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:26368 IMAGE:4816102"

/tissue_type="Brain, hippocampus"

/clone_lib="NIH_MGC_95"

/lab_host="DH10B"

/note="Sector: plasmidscript"

463..2121

/codon_start=1

/product="Similar to membrane-bound transcription factor

protease, site 1

/protein_id="AAH26330.1"

/db_xref="GI:20072868"

/translation="MKLVNIWLLLVLCOKKHLGRLEKSEFAKPCPCGSHLTK
 VFSSTVEYEVIVAFNGYFTAKARNSPFISSALXSSVDNRIIPNNPSSDPSDFE
 VIOKBKQGLLEDPNKRVTQKRVRSILKYAESDPTPCNETRWQKQWSSR
 PLRSLGSGFHWATGRHSSRLRLAIPROVAOTLOADVLMQWYTCANRVAVED
 TGLSEKHPKRVKERTNMTNERTLDGLGTFVAGVIAASRECQGFAPDAELHIFR
 VFTNNQSVAFNLDAFNALIKKIDVLNLSIGGPFDMHPFVDRKVELTANNIMVS
 AIGNDGLYTLNFPADQMDVIGGIDIFEDNIARFSRGMFTWLPGLGYGRMKPDIV
 TYAGYGVSGVKGCRALSGTSVAVAGVTLVSTVKRELNPASMKQALIASA
 RRLPGVNMFEQHGKLLRAYQILNSYKQPSLSPSYIDLTECPYMPYCSQPIYVG
 GMPTVYVNTLNGMGTGRIVDPQWQPYLPQNGDNIEAFVSYS"

778 a 707 c 797 g 744 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 2,47e-188 Length: 3026
 Score: 3069.00 Matches: 580
 Percent Similarity: 99.66% Conservative: 0
 Best Local Similarity: 99.66% Mismatches: 2
 Query Match: 54.64% Indels: 0
 DB: 9 Gaps: 0

US-09-830-837-6 (1-1052) x BC026330 (1-3026)

QY 1 MetLysLeuValAsnIleThrLeuLeuValLeuLeuCysGlyLysLysHis 20
 DB 463 ATGAAGCTTGTCAACATCGGCTGCTCTCGTGGTGTTCCTCTGGGAAGAAACAT 522
 QY 21 LeuGlyAspArgLeuGlyLysSerPheGluLysAlaProCysProGlyCysSerHis 40
 DB 523 CTGGCGACAGACTGGAAAGAAATCTTTTGAAGAGCCCATGCTGCTGCTGCCAC 582
 QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrIleValAlaPhe 60
 DB 583 CTGACTTTGAAGGTGGAATTCATCAACAGTTGTGGAATATGAATATATGTCCTTC 642
 QY 61 AsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSer 80
 DB 643 ATGGATACCTTACACCAAGCTAGAAATTCATTTATTTCAAGTCCCTGAGAGCACT 702
 QY 81 GluValAspAsnThrArgIleProArgAsnAsnProSerSerAspTyrProSerAsp 100
 DB 703 GAAGTGAACATTTGGAGAAATATACCTCGAAMCAATCCATCCAGTACTACCTAGTAT 762
 QY 101 PheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis 120
 DB 763 TTGAGGTGATTGAGATAAAGAAAAAGAAAGAGCGGGCTGCTTAACACTTGAATCAT 822

121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu 140
 DB 823 CCAACACATCAAAAGGGGTACGCCGCCCAAGAAAGTCTTTCGTTCCCTCAAGTATGCGAA 882
 QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpPheLysSerArg 160
 DB 883 TGTGACCCCAAGTACCTGCAATGAAACCGGCTGGAGCCAGAAAGTGGCAATCATCAGT 942
 QY 161 ProLeuArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis 180
 DB 943 CCCTCGCAAGAGCAGCCCTCTCCCTGGGCTCTGGCTTCTGCGTCTGACGGGAGCAT 1002
 QY 181 SerSerArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
 DB 1003 TCGAGCAGAGCGCTGCTGAGAGCCATCCCGCGCCAGGTTGCCAGACACTGCAGCAGAT 1062
 QY 201 ValLeuTrpGlnMetGlyThrGlyAlaAsnValArgValAlaValPheAsnThrGly 220
 DB 1063 GTGCTCTGGCAGATGGGATATACAGTCTTANTGTAAAGTGTGCTTTTGACACTGGG 1122
 QY 221 LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnThrAsnGlu 240
 DB 1123 CTGAGCGAGAGCATCCCACTCAAAATGTGAAGAGAGAACCACTGGACCAAGAG 1182
 QY 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
 DB 1183 CGAAGCTGGACATGGGTGGCCATGGCACATTCGTGGCAGGTGTATACCCAGCATG 1242
 QY 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
 DB 1243 AGGAGTGGCAAGGATTTCTCCAGATCAGAACTTCACATTTTCAGGGTCTTTTACCAAT 1302
 QY 281 AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLys 300
 DB 1303 AATCAGGTATCTTACGCATCTTGGTGTGGACGCTTCAACTATGCCATTTTAAAGAG 1362
 QY 301 IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisPropheValAsp 320
 DB 1363 ATCGACGTGTAAACCTCAGCATCGCGCGCCGACCTTCATGATCATCCGTTGTGTAC 1422
 QY 321 LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly 340
 DB 1423 AAGGTGTGGGAATTAACAGCTAACATCTATCATGTTTCTGCTATTGGCAATGACGGA 1482
 QY 341 ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyGly 360
 DB 1483 CCTCTTATGGCACTCTGATTAACCTCTGATCAATGGATGTGATGGAGTAGGCGGC 1542
 QY 361 IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThrTrpGluLeu 380
 DB 1543 ATTGACTTTGAAGATAACTACCGCCGCTTTCTTCAAGGGGAATGACTACTCTGGAGCTA 1602
 QY 381 ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly 400
 DB 1603 CCAGGAGGCTACGTCGCAATGAACCTGACATTTGTCACCTATGTTGCTGGCGTGGCGGGT 1662
 QY 401 SerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420
 DB 1663 TCTGCGGTGAAGGGGGTCCGCGCCCTCTCAGGGACCACTGTTGCTTCTCCAGTGGTT 1722
 QY 421 AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAla 440
 DB 1723 GCAGGTGCTGCACTTGTGTAGTGAACAGCTCCAGAGCGTGGAGTGGTGAATCCTGCC 1782
 QY 441 SerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMetPheGlu 460
 DB 1783 AGTATGAAGCAGCCCTGATCCGTCAGCCGAGGCTCCCGGGGTCAACATGTTTGA 1842
 QY 461 GlnGlyHisGlyLysLeuAspLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480
 DB 1843 CAAGGCCAGCGAAGCTCGATCTGTCAGAGCCTATCAGATCTCTCAACAGCTACAGCCA 1902


```

Qy 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
Db 1903 CAGCGAAGTTGAGCCCGACCTACATAGATCTGACTGAGTGTCCCTACATGTGGCCCTAC 1962
Qy 501 CysSerGlnProIleTyrTyrGlyGlyMetProThrValValAlaSerValThrIleLeuAsn 520
Db 1963 TGTCTCCAGCCCATCTACTATGAGGAGGAAATCCGACAGTGTATTAATGTCACCATCTCAAC 2022
Qy 521 GlyMetGlyValThrGlyArgIleValAlaSerProAspTrpGlnProTyrLeuProGln 540
Db 2023 GGCATGGGAGTCCAGAGAAATGTAGATAAGCCTGACTGGCAGCCCTATTGGCCACAG 2082
Qy 541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTrpSerGly 560
Db 2083 AAGGGAGACAACTGAAATGAGTGTCTCTCTACTCTAGCTCTTATGCGCTTGTGTCGGC 2142
Qy 561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln 580
Db 2143 TACTTGCCATCTCCATTTCTGTGACCAAGAAAGCGGCTTCTTGGGAAGCATTTGCTCAG 2202
Qy 581 GlyHis 582
Db 2203 GGCCAT 2208

RESULT 10
LOCUS AK091212
DEFINITION Homo sapiens cDNA FLJ33893 fis, clone CTONG2007681, highly similar
to Site-1 protease of sterol regulatory element binding proteins.
ACCESSION AK091212
VERSION AK091212.1 GI:21749529
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens tongue, tumor tissue cDNA to mRNA, clone_lib:CTONG2
clone:CTONG2007681.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1
AUTHORS Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamanoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3085)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1..3085
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CTONG2007681"
/tissue_type="tongue, tumor tissue"
/clone_lib="CTONG2"
/note="cloning vector: pME18SFL3"
BASE COUNT 760 a 763 c 790 g 752 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 6.19e-188 Length: 3085
Score: 3063.00 Matches: 569
Percent Similarity: 99.82% Conservativeness: 0
Best Local Similarity: 99.82% Mismatches: 1
Query Match: 54.53% Indels: 1
DB: 9 Gaps: 0

US-09-830-837-6 (1-1052) x AK091212 (1-3085)

Qy 483 SerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSer 502
Db 691 AGTTTGGAGCCCGACCTACATAGATCTGACTGAGTGTCCCTACATGTGGCCCTACTGTCTCC 750
Qy 503 GlnProIleTyrTyrGlyGlyMetProThrValValAlaSerValThrIleLeuAsnGlyMet 522
Db 751 CAGCCCATCTACTATGAGGAGGAAATCCGACAGTGTATTAATGTCACCATCTCAAGCGGATG 810
Qy 523 GlyValThrGlyArgIleValAlaSerProAspTrpGlnProTyrLeuProGlnAsnGly 542
Db 811 GGAGTCACAGAGAAAGTGTAGATAAGCCTGACTGGCAGCGCTATTATGGCCACAGACGGA 870
Qy 543 AspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTrpSerGlyTyrLeu 562
Db 871 GACAACATTTGAAGTTGCTTCTCTACTCTCGGTCTTATGGCCCTTGGTGGGCTACTGTG 930
Qy 563 AlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHis 582
Db 931 GCCATCTCCATTTCTGTGACCAAGAAAGCGGCTTCTTGGGAAGCATTTGTCAGGGCCAT 990
Qy 583 ValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThr 602
Db 991 GTCATGATCACTGTGGCTTCCCGACAGACAGACAGTCAAAAAATGTCGAGACACAGACT 1050
Qy 603 SerThrValLysLeuProIleLysValLysIleLeuProThrProProArgSerLysArg 622
Db 1051 TCACACAGT-AAGCTCCCATTAAGGTGAAGATAATCTCTACTCCCGCGGCAAGCAAGAGA 1109
Qy 623 ValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArgAspAsn 642
Db 1110 GTTCTCTGGGATCAGTACCACACACCTCCGCTATCCACTGGCTATTTCCCGAGGATAT 1169
Qy 643 LeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPhArg 662
Db 1170 TTAAGATGAAGAAATGACCTTTAGACTGGAATGGTGTATGATCAGTATCCATCCACCAATTTCCAGG 1229
Qy 663 AspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPhe 682
Db 1230 GATATGTACCAGCATCTGAGAAGCATGGGCTACTTTGTAGAGGTCTCTGGGGCCCCCTTC 1289
Qy 683 ThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGluTyr 702
Db 1290 ACCTGTTTGTATGCCAGTCACTAGTATGGCATTGCTGTATGGTGGACAGTGGAGGAGTAC 1349
Qy 703 PheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeuValIle 722
Db 1350 TTCCTCTGAAGAGATCGCAAGCTCCGAGGAGGAGTGGACAAACGGCTCTCGCTCGTCTATC 1409
Qy 723 PheSerAspTrpTrpAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThr 742
Db 1410 TTCAGTGACTGGGTACAACTCTGTATTATGAGAAAAGTGAAGTTTATGATGAAAACACA 1469
Qy 743 ArgGlnTrpTrpMetProAspThrGlyValAlaAsnIleProAlaLeuAsnGluLeuLeu 762
Db 1470 AGCAGTGGTGGATGCCGGATACCGGAGGAGCTTAACATCCAGCTCTGAATGAGTGTCTG 1529
Qy 763 SerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyPheThrLeuAlaAsn 782
Db 1530 TCTGTGTGGAACATGGGTTTCAGCGATGGCTGTATGAAGGGGAGTTTACCTTGGCCCAAC 1589
Qy 783 HisAspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValVal 802
Db 1590 CATGACATGTATTATTCGTGAGGTGCAGCATCGCAAGTTTCCAGAAAGTTCGAGTGTGCTG 1649

```

QY	803	IleThrGlnThrPhelYsAspGlnGlyLeuGluValLeuLysGlnGluThrAlaValVal	822
Db	1650	ATAACACAGAGCTTTCAAGGACCAAGGATGGAGTTTAAAGCAGGAAACAGCAGCTGTT	1709
QY	823	GluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyGlyArgIleVal	842
Db	1710	GAACACGTCCTTATGGAGCTTTATCAGATTCCAGCTGAGGCTGAGGCGCGGATGTA	1769
QY	843	LeuTyrGlyAspSerAsnCysLeuAspSerHisArgGlnLysAspCysPheIlePleu	862
Db	1770	CATGATGGGACCTCAATGCTGGATGACAGTACCACAGAGAGGACTGCTTTGGCTT	1829
QY	863	LeuAspAlaLeuLeuGlnThrSerTyrGlyValThrProProSerLeuSerHisSer	882
Db	1830	CTGATGCTTCTCCAGTACATCTGATGGGTGACAGCCGCTACCTCAGTCACTCT	1889
QY	883	GlyAsnArgGlnArgProSerGlyAlaGlySerValThrProGluArgMetGluGly	902
Db	1890	GGAAACCCGAGCCCTCCAGTGGAGCAGCTCAGTCACTCCAGAGAGATGGAGGA	1949
QY	903	AsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLysProArg	922
Db	1950	AACCATCTTCATCGTACTCCAAAGTCTGGAGGCCCAFTTGGAGACCCCAAACTCG	2009
QY	923	ProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGluThrAla	942
Db	2010	CCTCTACAGCCTGTCCAGCCTGTCTTGGGCCCAAGCCACAGCCTTTAAACGAGAGCG	2069
QY	943	ProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysValLeu	962
Db	2070	CCAGTAACTTGGAAACATCAGAACTACTTCCATTTGACCTGGACAGAGTGGTGA	2129
QY	963	ProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSerGlyAla	982
Db	2130	CCCACTTTCGATCGATCGCTCAAGTGAAGGCCCTTGTCCCTTGGAGAGAGCGCGC	2189
QY	983	TrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGlnThrIle	1002
Db	2190	TGGACATCTCTGGAGGATCATGCTGGCGGCTACAAACGAGAGTGGCGGACCAT	2249
QY	1003	ProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGlnIleAsn	1022
Db	2250	CCGTCTTTCGCTTCGAGGAGCCATGGTGTCTGGCCTTCTTGTGTACAAATCAAC	2309
QY	1023	LysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeuMetGln	1042
Db	2310	ARGCCAGAGAGCGGCGGAGCGAGGAGCCGAGGTCGAGCGCCGCGGACGTCATG	2369
QY	1043	GlnValHisProProLysThrProSerVal	1052
Db	2370	CAGGTTCAACCCGCAAGACCCCTTCGGTG	2399
RESULT 11			
LOCUS	AF441758		
DEFINITION	Drosophila melanogaster serine endopeptidase mRNA, complete cds.		
ACCESSION	AF441758		
VERSION	AF441758.1		
KEYWORDS	1 (bases 1 to 3136)		
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster.		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	1 (bases 1 to 3136)		
TITLE	See Miller, A.C., Dobrosotskaya, I., Goldstein, J.L., Ho, Y.K., Brown, M.S. and Rawson, R.B.		
JOURNAL	Drosophila melanogaster serine endopeptidase mRNA, complete cds.		
MEDLINE	Dev. Cell 2 (2), 229-238 (2002)		
PUBMED	21822137		
REFERENCE	2 (bases 1 to 3138)		
AUTHORS	See Miller, A.C., Dobrosotskaya, I., Goldstein, J.L., Brown, M.S. and Rawson, R.B.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-NOV-2001) Molecular Genetics, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-9046, USA		
FEATURES	Location/Qualifiers		
source	1..3138		
	/organism="Drosophila melanogaster"		
	/strain="Oregon R"		
	/db_xref="taxon:7227"		
	/chromosome="3"		
	/map="3L"		
	/dev_stage="embryo"		
	58..3096		
	/note="dsIP; similar to CP22149"		
	/codon_start=1		
	/product="serine endopeptidase"		
	/protein_id="AAM20922.1"		
	/db_xref="GI:20378355"		
	/translation="MNVTPLFIISACISLDAFKTAVVPNEFIVHFHSKFAPVRESY IAKLLGSNTVNRIVPRLNLAQPSDFDLKRCVGVSESEFIEBRLQTHPSKAV VPORSRRILNYDAYSLNTYIHRHPQVLRNPNDRHROLCSVLHNLKLGITG KGVVAIFDTGLTKNHFNRVKTNTNEKSLDDRVSHGTFVAGVIASSRECLGFA PDADLIFKFTNSOVSYTFNDAENYAIYKINLNLISLGGPDMSPFVKEVLEL SANNVIMISAAGNDQPLXGTLNPGQSDVGVGGIOPDDKIAKFSRGTWTWELPLG YRNGLDIVTGSQVSGDVEKGRISCTSVSSPVVAGAAALLIGAFQKIDYINPA SLKQVLEGRGLRPHYNFEGAGKLNLSKMLLSVYKPKITLIPALDFTONYMP YSOPLITGSSVALIANVILNGISVISHVIGPKIPROPENOGLOVSAQVSPVMP WTGMSVFTAVKRGNEFEGVCKGSIITLVLESEKQTNETHVTEVDFPLTIKVPKPP RNRKILMDQTHSLRYPRIIPRODLKVLQDLPDNRADHILHFNROMYTHLRNYYVID VLEPFFTCNASDIALLVDPFGDEINALQENYIKGLNLYVFGVGYNTYMK KIKFFPNQDMTDTGGANIPALNDLLKPGIAPDFGVGSHFKGLDSMYSVAGA TIYKFPNPDGIIVGTKLNDQGLSILNSKTPSKVAKLDVIFGMFTANASQSNEL VVAENSLAEAPIDTSTFKNRVLLTKORSISPAKNNHETNEGRIVAGSINCL DSHLEKACTWLLITFDLFAINSLKSLONUNRIFETKHEPLRPLRISQSIKSR SDNCEQEKWLAPTKONNAERKSSIDVTILENEHEHINILNKLGEETAKLGONN DYLTGQMSADSLTPFIILISLVIMEVLEFLKRRFIL"		
BASE COUNT	981 a 662 c 656 g 839 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	37e-144	Length:	3138
Score:	2383.50	Matches:	478
Percent Similarity:	59.83%	Conservative:	167
Best Local Similarity:	44.34%	Mismatches:	272
Query Match:	42.43%	Indels:	161
Db:	3	Gaps:	17
US-09-830-837-6 (1-1052) x AF441758 (1-3138)			
QY	47	PheSerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAla	66
Db	112	TTTAAACAGCGGTGTTCCTCAAGAGTCACTCCTTCCATCAAAATACTTGCC	171
QY	67	LysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArg	86
Db	172	CCGGTCCGAGAACTCTACATCGCAGCAAACTTCTGGTTCAACGATACGACGAGGA	231
QY	87	IleIleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIle	106
Db	232	ATTGTTCCTCCCTTAAATTTGGCTTGGCAATATCAAGTATTTGATATCTTACAGTT	291
QY	107	LysGluLysGlnLysAlaGly-----LeuLeuThrLeuGluAspHisProAsn	122
Db	292	TGGCAGCGTTTGAATCATCATCATCATATATATAAGAGCGCTTCAGACTCACCATCA	351
QY	123	IleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyr---AlaGluSer	141
Db	352	GTAAGGAGTAGTTCCCGCAGCGGAGGAGTCTCTAAACTATGACGCTTATAGC	411
QY	142	AspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerArgPro	161

Db 412 AACCTAACG----- 420
Qy 162 LeuArgArgAlaSerLeuSerLeuGlySerGlyPheThrPheHisAlaThrGlyArgHisSer 181
Db 421 -----TATATTCAC-----CGGCATCCC 438
Qy 182 SerArgArgLeuLeuArgAlaIlePro-----ArgGlnValAlaGlnThr 196
Db 439 CAAGGAGTGTAAAGGAACAGAAACCAACACGATCGCCACCGACAAATGTGTCGGTA 498
Qy 197 LeuGlnAlaaspValLeuTrpGlnMetGlyThrGlyAlaAsnValArgValAlaVal 216
Db 499 CTCACGCCAACATCCCTTTGGAAGCTGGGTATCATCAGGCAAGGAGTAAAGTGGCCATT 558
Qy 217 PheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsn 236
Db 559 TTCGACACTGGCCTAACCAAAACCATCCACACTTCGAAATGTAAAGGAACGAACAAC 618
Qy 237 TrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyVal 256
Db 619 TGGACGAATGAAAGTCACTTGACGACAGAGTCACTGTCATGCGACCTTCGTCGCGGGGTA 678
Qy 257 IleAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArg 276
Db 679 ATCGCTTCTCCAGGGAATGCTAGGCTTCGCTCCCGACCGCGATCTTACATATTTAAA 738
Qy 277 ValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAla 296
Db 739 GTTTTACGAACCTCCCAAGTTCTTACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 798
Qy 297 IleLeuLysLysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHis 316
Db 799 ATATATAGGAATAAACAATCTCAACCTTAGCATAGCATTTGGGGTCCCGACTTTATGGACTCG 858
Qy 317 ProPheValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIle 336
Db 859 CCGTTCGTGAAAGGTGTGGAACTGTCGCTTAATAATGTCAATATGATATGGCAGCA 918
Qy 337 GlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIle 356
Db 919 GGAATGATGTCCTTGTACGGCACGCTAAACAATCTTGGCGATCAGAGCGATGTAGTT 978
Qy 357 GlyValGlyIleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThr 376
Db 979 GCGCTGGTGCAATCACTAGCTAGGCTAGGAGTATGGAGTCCCAAGTTTACTTCAGAGGAATGACA 1038
Qy 377 ThrTrpGluLeuProGlyGlyGlyArgMetLysProAspIleValThrTyrGlyAla 396
Db 1039 ACGTGGGAACCTCCCTTAGCTAGGCTAGGAGTATGGAGTCCGATATGTCAGTACGGAAGT 1098
Qy 397 GlyValArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1099 CAAGTGAAGGACAGTATGTCGCAAGGGGTGACAGCAGCTCTCTGGAACATCCGTGTC 1158
Qy 417 SerProValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGlu 435
Db 1159 TCTCCAGTTGTGAGGGGTGCTGCACTGCTGCTATAAGCGGTGCTATTCAGAAATCGAC 1218
Qy 436 LeuValAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGly 455
Db 1219 TACATAACCCAGCATCTTTAAGCAGTACTCAATGAAGTCCGCGAGAACTCCCGCAT 1278
Qy 456 ValAsnMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeu 475
Db 1279 TATAACATGTTGAGCAGGGAGCTGGAAACTGAATTTGCTGAAGAGTATGCACTATTG 1338
Qy 476 AsnSerTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysPro 495
Db 1339 CTGTATACAAACCAAGATAACCTTATTCGCGCATACCTTGACTTCAACCCAA---AAC 1395
Qy 496 TyrMetTrpProTyrCysSerGlnProIleTyrGlyGlyMetProThrValValAsn 515
Db 1396 TATATGGGCTTATAGCTCCCAACCTCTGTACTTGAAGCTCCGCTCGCTATTGTCAAC 1455

Qy 516 ValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGln 535
Db 1456 GTTACCACTACTCAATGATCTCTGTCACAGTCATATAGTTGGCATCCCTTAATGGATT 1515
Qy 536 ProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeu 555
Db 1516 CCCGATTTCGAAAACCAAGTCAGTTCTTCAAGTATCTGCACAAGTTTCCGCTATCGTT 1575
Qy 556 TrpProTrpSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrp 575
Db 1576 TGGCCGCGGAGCGGTCAGTTTATCTGCTGTAATAAAGGAAGAGAGAACTTT 1635
Qy 576 GluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThrGluSer 595
Db 1636 GAAGTCTTTGTAAGGAAGTATCACCTAGTTTGGAAAGCTTTAAACAGACACCACCAAC 1695
Qy 596 LysAsnGlyAlaGluGlnThrSerThrValLysLeuProIleLysValLysIleIlePro 615
Db 1696 GAAACT-----CATGTTACAGAAGTCGACTTCTCTTAAACAATAAAGGTTACTCCA 1746
Qy 616 ThrProArgSerLysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProPro 635
Db 1747 AAACCCGCAAAACAAAGAGATTTATGGATCAGTACCACACCTTAGGATATCCACCG 1806
Qy 636 GlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAsp 655
Db 1807 CGCTATATCCACGAGATGATCAAAAGTTAAACTAGATCCTCTGGACTGGAGGCGACAC 1866
Qy 656 HistIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheVal 675
Db 1867 CATATACACAAACTTTAGGACATGATATACACATTTAGCAATTTGGCTACTACATT 1926
Qy 676 GluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMet 695
Db 1927 GATGTTTGGCAGAACCTTCACTGCTTCAATGCCCTCGGATTTAGCGGCTTATGATT 1986
Qy 696 ValAspSerGluGluGluTyrPheProGluGluIleAlaLysLeuArgAspVal--- 714
Db 1987 GTTGACCTGAGAGAGGTTTGGCGACGAGGAAATAAAGCTTTACAGGAAACGCTGTAT 2046
Qy 715 AspAsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThrSerValMetArgLys 734
Db 2047 AAAAGAGGCTGAATGCTGCTGATTCGGAGACTGGTATACACACTGTGATGAAATAA 2106
Qy 735 ValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGlyGlyAlaAsn 754
Db 2107 ATTAATCTTTGACGAGACACCCGACATGGTGGACACCCGACACTGGTGGCGCAAT 2166
Qy 755 IleProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyr 774
Db 2167 ATTCAGCTTGAATGATTTATTGAAGCATTTGGAATTTGCTTTGGCGATTTGTCGGT 2226
Qy 775 GluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAla 794
Db 2227 GAGGACATTTCAAACTGGCGCACCATTCACTACTACTAGTGTAGTGAGCCACCAATGTT 2286
Qy 795 LysPheProGluAsp-----GlyValValIleThrGlnThrPheLysAspGlnGlyLeu 812
Db 2287 AGTTTCAATGAATCCAGGAGATATTATAGTGGGCACAAACCTGAATGACCAAGGACTT 2346
Qy 813 GluValLeuLysGlnGluThr-----AlaValValGluAsnValProIleLeuGly 829
Db 2347 TCGATTATTATTTAAACACACCCAGCAAGTAGCAAAACTAGATGTACCTATTTTGGT 2406
Qy 830 LeuTyrGlnIleProAla----- 835
Db 2407 ATGTTCCAAACCAAGGCGAAGTATTCAAAGCAACAGGAAATCGTGGTCAATGCGGAA 2466
Qy 835 ----- 835
Db 2467 AGCAATTTGGCAGAGGCTATATCCACAGATTACTCCACATTTAAGAACCGGGTTTCTCA 2526

QY	836	-----	Glucly 837
Db	2527	CTGCGAAGCAAGCAAGTATCATGTTTCGAAAGCAATATCATGAAGTAAAGTAAAT	2586
QY	838	GlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnIys	837
Db	2587	GAAGACATATGCGCTATGATGGGAGCTATCAATGCTCCAGCTCCAGCTGGAGAA	2646
QY	858	AspCysPheThrLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrPro	877
Db	2647	GCTTGCTACTGGCTGCTAATACGCTTTTATGATTTTGAATATGAATAAAT	2691
QY	878	SerLeuSerHisSerGlyAsnArgGlnArgProSerGlyAlaGlySerValThrPro	897
Db	2692	-----TCGCAC-----	2697
QY	898	GluArgMetGluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGly	917
Db	2698	-----AAATCAAGTTTATTCGACAAATCAATTCGTAATATCACTGAATTCACAAATTA	2748
QY	918	AspProLysProArgProLeuProAla-----	926
Db	2749	GAGAGAGCACCATTACCCCTTAGGATATCCGAAGTATTAATAAATCTGTTACAGGAC	2808
QY	927	-----CysProArgLeuSerTyrAlaLysProGlnProLeuAsnGlnThrAlaProSer	944
Db	2809	ATAATTTGTGACAAATTAATGCTGTCGACCGACGACGACAAATATACGCC-----	2859
QY	945	AsnLeuTyrLysHisGlnLysLeuLeuSerLysLeuAspLysValValLeuProAsn	964
Db	2860	-----GAGGAAGGAATCTCTTAATAGACGTPAACCATATCTGGAAT	2904
QY	965	PheArgSerAsnArgProGlnValArgProLeuSerProGlyGlySerGlyAlaThrPasp	984
Db	2905	GAAGACACGAGATAAATTTATCAAAATATTTATGCTGGAGGATCGCAAACTAGGG	2964
QY	985	IleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGlnThrIleProVal	1004
Db	2965	CAAAACAATGATTTATTTACAGAGGAATCAATCGCGGATGCTTAATGACTCCATATTA	3024
QY	1005	PheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValGlnIleAsnIysAla	1024
Db	3025	TTCTTAATTTATTAAGCCTAATGTTATCATGTTATTTGTTATTTTAAAGCGTCGA	3084
QY	1025	-----LysSerArgProLysArgLys 1032	
Db	3085	TTCAATTTGTTAAATAATCTCCACTTTTAAAAAACGGTAAAAA	3138
RESULT 12			
AC010701			
DEFINITION	AC010701	174920 bp DNA linear INV 27-DEC-2001	
AC010701		Institute Drosophila BAC Library) complete sequence.	
AC010701.3	GI:15412430		
KEYWORDS			
NG			
SOURCE			
ORGANISM			
Drosophila melanogaster.			
Drosophila melanogaster.			
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
Phyloidea; Drosophilidae; Drosophila.			
1 (bases 1 to 174920)			
Muzny, D., Scherer, S., Adams, M.D., Holt, R.A., Evans, C.A.,			
Gocayne, J.D., Tabor, P., Williamson, A., Homs, F.H.,			
Dugan-Rocha, S.D., Sodergren, E.H., Hodgson, A., Chen, R.C.,			
Ayele, M.A., Scott, G.S., Worley, K.W., Amaratunga, P.G., Brandon, R.C.,			
Rogers, Y., An, H., Baldwin, D., Beeson, K.Y., Brown, M., Buhay, C.,			
Busan, D.A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K.,			
Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E.,			
Draper, H., Emery-Cohen, A., Ferrera, S., Garg, N.D.S., Houck, J.,			
Hoskins, D., Howland, T.J., Hume, J., Ibegwam, C., Jalali, M., Kovar, C.,			
Liu, W., Mattei, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B.,			
Nelson, K.A., Nadesa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V.,			
Scherer, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M.,			
Xiang, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O.,			
Wheeler, D., Weinstock, G., Gibbs, R. and Venter, J.C.			
Direct Submission			
Unpublished			
2 (bases 1 to 174920)			
Worley, K.C.			
Direct Submission			
Submitted (18-SEP-1999)			
Human Genome Sequencing Center, Department			
of Molecular and Human Genetics, Baylor College of Medicine, One			
Baylor Plaza, Houston, TX 77030, USA			
3 (bases 1 to 174920)			
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,			
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbara, J.,			
Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,			
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,			
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carton, T.F.,			
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,			
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,			
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,			
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,			
Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H.,			
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,			
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,			
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,			
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,			
Hamilton, K., Harris, K., Harris, K., Hart, M., Havlak, P., Hawes, A.,			
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,			
Joudan, S., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,			
Kovar, C., Kratovic, J., Kureshi, A., Land, U., King, L., Korvah, J.,			
Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,			
Loeughele, H., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,			
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,			
Martinez, E., Massey, E., Mahoney, B., McLeod, M.P., Meador, M.,			
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Monabhat, K.,			
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,			
Nguyen, A., Okunou, G., Oragunye, N., Nickerson, E., Nwokenkwo, S.,			
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,			
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,			
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,			
Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,			
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,			
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,			
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,			
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,			
Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, S.,			
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,			
Worley, K. and Gibbs, R.			
Direct Submission			
Submitted (01-SEP-2001)			
Human Genome Sequencing Center, Department			
of Molecular and Human Genetics, Baylor College of Medicine, One			
Baylor Plaza Houston,			
(bases 1 to 174920)			
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,			
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbara, J.,			
Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,			
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,			
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carton, T.F.,			
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,			
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,			
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,			
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,			
Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H.,			
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,			
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,			
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,			
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,			
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,			
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,			
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,			


```
QY 277 lphetrAsn----- 280
Db 53985 TTTTACGAACTCCCAAGTGGAGCAACGCTAGAACCTGAAATTAACCTAAGCAATTAATAG 54044
QY 281 -----Asn-GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIle 298
Db 54045 CCTACATCGCGCAGGTTTCTTACCTTCTGGTTCTCGATGCAATTCACATGACCGCATAT 54104
QY 298 eulysylleAspValLeuAsnLeuSerileGlyGlyProAspPheMetAspHisProp 318
Db 54105 ATAGGAAATAACATCTCAACCTTAGCATTTGGGGTCCCGACTTTTAGCATCTGCCCT 54164
QY 318 heValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyA 338
Db 54165 TCGTTGAAAGGTGTGGAACTGTGCGCTAATATGCTAATATGATATCGCGACAGGAA 54224
QY 338 snAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyV 358
Db 54225 ATGATGGTCCCTGTACGGCAGCTTAACATCTCGGCATCGAGCGATGTAGTTGGCG 54284
QY 358 alGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrThrT 378
Db 54285 TTGGTGGATTCAGTTTGATGATAAAATCGCAAGTTTAGTTGCGAGGATGCAACGT 54344
QY 378 rpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyV 398
Db 54345 GGAATCTCCCTTAGGCTACGACGATGGACTTCGATATTCTACGTACGGAAGTCAAG 54404
QY 398 alArgGlySerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerP 418
Db 54405 TGAAGGAGGATGTGTCGCGCAAGGGTGCAGACGACTCTCTGGACATCGCTGCTCTC 54464
QY 418 roValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGluLeuV 437
Db 54465 CAGTTGTTCAGGGGTGTGCTGCTTAAAGCGGTGCATTCAGAAATACGACTACA 54524
QY 437 alAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValA 457
Db 54525 TAAACCCAGCATCTTAAGCAGGTACTCTTAAGGTGCGGAGAACTGCGCATATA 54584
QY 457 snMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnS 477
Db 54585 ACATGTTTGACAGGAGCTGGAAACTGAATTTGCTGAAGAGTATGCAGTATCTCT 54644
QY 477 erTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrM 497
Db 54645 CATACAAACCAAGATAACCTTATTCGGCATACCTTCCTCACCACAA---AACTATA 54701
QY 497 etTrpProTyrCysSerGlnProIleTyrGlyGlyMetProThrValValAsnValT 517
Db 54702 TGTGGCTTATAGCTCCCACTCTGTACTATGGAAGTCCGCTCGCTATGCAACAGTTA 54761
QY 517 hrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProT 537
Db 54762 CCATACTCAATGGTATCTCTGTCAAGATCATATAGTTGGCATCCCTCAATGATTCCCG 54821
QY 537 yrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpP 557
Db 54822 ATTTCGAAACCAAGGTCAAGTTCTTCAAGTATCTGCACAAAGTTTCGCTTATGCTG 54881
QY 557 roTrpSerGlyTyrLeuAlaIleSerIle----- 566
Db 54882 CGTGACCGGTGTGATGTCAGTTTATTTGTTGAGTATGAAATAGCAACGCTATATAA 54941
QY 566 ----- 566
Db 54942 AGGATATATGGGTACATATTGATATGAAAGTCTAAGCAATTCATACCAACAATTT 55001
QY 567 -----Ser-ValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisVal 583
Db 55002 TGTTTACACGCTGTAAAGAGGAGGAGAAACTTTGAAGGTGTTTGAAGGAGTATC 55061
QY 584 MetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThrSer 603
```

```
Db 55062 ACCTAGTTTGGAAAGCTTTAAACAGACACCACCAACAACT-----CATGTTACA 55112
QY 604 ThrValLysLeuProIleLysValLysIleIleProThrProProArgSerLysArgVal 623
Db 55113 GAAGTCGACTTTCCTTTAAACAATAAGGTACTTCCAACACCGCAAGAAACAGAGGATT 55172
QY 624 LeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArgAspAsnLeu 643
Db 55173 TTATGGGATCAGTACCACACGCTTAAGGTATCCACGGCTATATTCACAGAGATGATCTC 55232
QY 644 ArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgAsp 663
Db 55233 AAGTTAAACTAGATCTCTGGACTGGAGGCGACACCATATACACAAACTTTAGGGAC 55292
QY 664 MetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThr 683
Db 55293 ATGTATACACATTTACGAATGTTGCTACTACATGATGTTTGGGAGAACCTTCACC 55352
QY 684 CysPheAspAlaSerGlnTyrGlyThrLeuMetValAspSerGluGluTyrPhe 703
Db 55353 TGTCTCAATGCTCGGATGATGCGGTTTATGATGTTGACCTGCTGAGAGGTTTGGC 55412
QY 704 ProGluGluIleAlaLysLeuArgArgAspVal---AspAsnGlyLeuSerLeuValIle 722
Db 55413 GACGAGGAATAAACGCTTTACAGGAAACGCTGTATAAAGAGCTTGAATCTTGTCTGTA 55472
QY 723 PheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThr 742
Db 55473 TTCGAGACTGGTATACACACCTGTGTATGAAATAAATAATCTTTCAGCCCTGAATTTAT 55532
QY 743 ArgGlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLeu 762
Db 55533 CGACAATGCTGGACACCGCACCTGTGGCGCAATATTCAGCCCTGAATGATTTATTTG 55592
QY 763 SerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsn 782
Db 55593 AAGCCATTGGATTTGTCGATTTTGGCGATTTTGGTGGAGGACATTTCAAACTGGCGAC 55652
QY 783 HisAspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluVal-----Gly 800
Db 55653 CATTCATGTACTATGCTAGTGGAGCCACATTTGAAGTTTCCATGATGATCCAGGAT 55712
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuVal----- 814
Db 55713 ATATAGTGGCGCACAAACTGAATGCCAGAGCTTTCGGTAATTAATATATATATCTTT 55772
QY 815 -----LeuLysGlnGluThr 819
Db 55773 TTAGTTTCCACAGAAATACAACTTTCCTTTTCCAGATTAATATTAATAACACCC-AGC 55831
QY 820 AlaValValGlu---AsnValProIleLeuGlyLeuTyrGlnIleProAla----- 835
Db 55832 AAGGTAGCAAACTAGATGTACCTATTTTGGTATGTATTCACAAACGAGCGAACAGTATT 55891
QY 835 ----- 835
Db 55892 CAAAGCAACGAGGAAATCGTGTCAATCGGAAAGCAATTTTGGCAGAGGCTATACCCACA 55951
QY 835 ----- 835
Db 55952 GATTACTCCACATTTAAGAACCGGTTTGTCTACTGCGAAACGAGCAACGAGTATCAGT 56011
QY 836 -----GluGlyGlyValArgIleValLeuTyrGlyAsp 846
Db 56012 TTTGCAAAAGCAATAATCATGAAACTAAGAATCAAGAGCTATGTCGCTATATGGGAC 56071
QY 847 SerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeu 866
Db 56072 TCCAACTGCCCTCAGCTCCACGATCTGGAGAGCGCTGCTACTGCTGCTGCTGCTGCTGCT 56131
QY 867 LeuGlnTyrThrSerTyrGlyValThrProProSerLeuSerHisSerGlyAsnArgGln 886
```

Db 56132 TTAGATTTCGAATAAC-----TCGCAC----- 56155

QY 887 ArgProProSerGlyAlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHis 906
||| |||
||| |||

Db 56156 -----AAATCAAGTTTATTGCAG 56173

QY 907 ArgTySerLysValLeuGluAlaHisLeuGlyAspProLysProArgProLeuProAla 926
||| |||
||| |||

Db 56174 AATCTAAATCGTATACTGAATTCACAAATTAGAGAGACCACTACCCCTAGGATA 56233

QY 927 -----CysProArgLeuSerTrrPala 933
||| |||
||| |||

Db 56234 TCACAAGATTTATAAAATCTCGTTCACAGGACAATAATTGTGAACAATTTAAGTGGCTT 56293

QY 934 LysProGlnProLeuAsnGluThr-AlaProSerAsnLeu----- 946
||| |||
||| |||

Db 56294 GCACCCAGCAGCAAAATACGCCGAGGAAGAAATCTCTATATATAGACGTAAACATA 56353

QY 947 -TrpLysHisGlnLysLeuSerLeuAspLeuAspLysValValLeuProAsnPheAr 966
||| |||
||| |||

Db 56354 CTGGAATAATGAAGGTATTTATATACATTTAGTTCAACTACAAAGTCTCATATTTGTTGT 56413

QY 966 gSerAsnArgPro-----GlnValArgProLeuSerProGlyGluSerGlyAlaTrpAs 984
||| |||
||| |||

Db 56414 TTCAGAACACGAGATAAATTTATCAAAATTTATTGGGTGAGGAGATCGCAAACTAGG 56473

QY 984 pileProGlyGlyLeuMetProGlyArgTyAsnGlnGluValGlyGlnThrileProVa 1004
||| |||
||| |||

Db 56474 GCAAACATGATTATTATACAGGAATGCATCCGGGATAGTCTAATGACTCCAATAT 56533

QY 1004 lPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValGlnIleAsnLys 1023
||| |||
||| |||

Db 56534 ATTCTTAATATATTAAGCTAATTTGTTATCATGTTATTGTTTATTTTAAAGCGT 56591

RESULT 13

AE003595/c

LOCUS 294308 bp DNA linear INV 04-OCT-2000

DEFINITION Drosophila melanogaster genomic scaffold 14200001386036 section 5 of 9, complete sequence.

ACCESSION AE003595

VERSION AE002647

KEYWORDS AE003595.2 GI:10726941

SOURCE HTG.

ORGANISM Drosophila melanogaster.

REFERENCE 1 (bases 1 to 294308)

AUTHORS Adams,M.D., Celnikier,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.F., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahake,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S., Fleischmann,W., Fosler,C., Gabriellian,A.E., Garb,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matteli,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shner,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,C.J.

The genome sequence of *Drosophila melanogaster*

Science 287 (5461), 2185-2195 (2000)

20196006

MEDLINE 10731132

REFERENCE 2 (bases 1 to 294308)

AUTHORS Adams,M.D., Celnikier,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.

TITLE Direct Submission

JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT On Oct 9, 2000 this sequence version replaced gi:7296463.

FEATURES

source Location/Qualifiers

1. .294308

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="3L"

/complement(<322. .>837)

/gene="CG14574"

/evidence=not_experimental

/map="78E2-78E4"

/db_xref="FLYBASE:FBan0014574"

/db_xref="FLYBASE:FBgn0037104"

/complement(<322. .>837)

/gene="CG14574"

/product="CT34306"

/evidence=not_experimental

/db_xref="FLYBASE:FBan0014574"

/db_xref="FLYBASE:FBgn0037104"

/complement(322. .837)

/gene="CG14574"

/note="CG14574 gene product"

/codon_start=1

/evidence=not_experimental

/protein_id="AAF51750.1"

/db_xref="GI:7296464"

/db_xref="FLYBASE:FBan0014574"

/db_xref="FLYBASE:FBgn0037104"

/translation="MQPTLECSILGSWASEIPNVFDLNGGTTDAGGAAAAALSCS RYPRINATSPCOTQRCRGPAPPAHLQPHLLLVRGEGDGPGRCTTWRDWSGRL DSGNGKDKRGCGGTSTERACFCQCDISGLPTGHCNPNRDLDKAPGPGTQSGHSTH NKPSPSQKAR"

<27608. .>32838

/gene="AP-2"

/note="CG7807"

/map="78E4-78E5"

/db_xref="FLYBASE:FBan0007807"

/db_xref="FLYBASE:FBgn0023417"

join(27608. .27865,28904. .29134,29204. .29371,30628. .30835, 31480. .31669,31730. .31994,32098. .32265,32329. .32838)

/gene="AP-2"

/product="CT3720"

/note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."

/db_xref="FLYBASE:FBan0007807"

/db_xref="FLYBASE:FBgn0023417"

join(27821. .27865,28904. .29134,29204. .29371,30628. .30835, 31480. .31669,31730. .31994,32098. .32265,32329. .32439)

/gene="AP-2"

/note="AP-2 gene product; Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript"

gene	/codon_start=1 /protein_id="AAF51751.1" /db_xref="GI:7296465" /db_xref="FLYBASE:FBan0007807" /db_xref="FLYBASE:FBgn0023417" /translation="MTLAYVDMPAVIOVQERLSHGSHGLSSGSAAYTHSGGHTGRS GPATGSHGSHSTHGAATMHQSLQSDQFPYPPFPHHSTQSPQOQNHGALLEYLGT RSDYGAISAGAAGHCTRLGHEHESLALHOALQNAVDDVQAPALQNDNVAFMSDPLIKS MKSCKAGNIGSGSPSVFCAVGPLSLISLSTKYKVTIAEVRRLSPPECLNALSILG GVLRAKSNKGRLLREKLEKIGLNLPGARRKAAANTLLTSVVEGATHLAKDFHVC ETEPAROLAIEYVHOTEQSDYSRRKELILHSQOITKELMQLISQDRTTHFGRSOH LLEPSMORHLTHSLTHGRGSPAIMAVLHAFQIFLNSLYLEKLYPSNGGGMVSS LDKSKIDNEKK" complement(<33461..>36792) /gene="CG7169" /map="78B4-78E5" /db_xref="FLYBASE:FBan0007169" /db_xref="FLYBASE:FBgn0037105" complement(join(<33461..34108,34165..34904,35005..35857, 35997..36455,36634..>36792)) /gene="CG7169" /product="CG22149" /db_xref="FLYBASE:FBan0007169" /db_xref="FLYBASE:FBgn0037105" complement(join(<33461..34108,34165..34904,35005..35857, 35997..36455,36634..>36792)) /gene="CG7169" /note="CG7169 gene product" /codon_start=1 /protein_id="AAF51752.2" /db_xref="GI:10726942" /db_xref="FLYBASE:FBan0007169" /db_xref="FLYBASE:FBgn0037105" /translation="MNVFTFLFIISAICSDLAFTAVVPNEFIVHFSKYFAPVRESY IAAKLGSNNWNRIVPRNLNWOYPSDEFILRVCDSGESSEFIERLQTHPSVKAY VQPSVRILNYDAYSLNLYTHRPQGVLRNPNRNDHRQLCSVLHANNILWKLGIIG KGVAIFDTGLTKNHPFRNKRERNTWNEKSLDDRVSHGTGAGVSYTSWFLDAFN YAIRKINILNLSIGGDPFMSPEVEKLELSANNVIMISAAGNDGPLYGLTILNPPGO SDVVGVOFDDKIAKFSRSGMTTWELPGLYGRMLDIVYQSGVSDVRKGCRL LKSMSLLSYKPKITLIPALDFTQNYMPPYSOPLYGSSVAIANVTILNGISVTS HVGIPKWDPENQGVQVSPVIVPWTGMSVFAVKKEGNEGVCVGSIT LVLSFKQTNETHYTEVDPLTKVTPKPNKRILWDOYHSLRYPYPRIPRODLKAV KLPLDMRADHILHTFRDMTHLRNNGYIIDVLRPEPTCFNADSYGALLIVDPERGFG DEELNALQENYKRLGNVVFQWNTYVKKIKFFEDTROWMTDGTGGMANIPALND LLKPGIAGDFVGRGFKLGDHSMYASGATVIFKPMNPGDILIVGTLKNDGILSIIN SKTSPKVALDVPFQFQKANSIOSNEEIVVNAESIAFAIPTDYSTFKNRVILLR TKQBSIFSAKSNHETKNEGLIAYGDSNCLDSTHLKACYLLITFLDFAINSHKSS LLQNLNRIEFHKLERAPLRISQSLIKRSQDNCKQFKWLAPTQKQNAERKSSI IDVILENKGIVLHVLQSLIFVCFRTDRKFNQKFIG" /db_xref="FLYBASE:FBan0011307" /db_xref="FLYBASE:FBgn0037106" join(37155..37175,37839..38112,38173..38460,38710..39358) /gene="CG11307" /product="Ct43349" /db_xref="FLYBASE:FBan0011307" /db_xref="FLYBASE:FBgn0037106" join(37159..37367,37839..38112,38173..38460,38710..39358) /gene="CG11307" /product="Ct31547" /db_xref="FLYBASE:FBan0011307" /db_xref="FLYBASE:FBgn0037106" join(37853..38112,38173..38460,38710..38992) /gene="CG11307" /note="CG11307 gene product [alt 1]" /codon_start=1 /protein_id="AAG22183.1" /db_xref="GI:10726943" /db_xref="FLYBASE:FBan0011307"		CDS	/db_xref="FLYBASE:FBgn0037106" /translation="MEELCNSSGHSATSSSGNSSISAKTALSCSAATNYLSALNN LCTAGSKLAHSIAVLEQNSISERLENNYTYLJNSWDLARATTVATGVTKTHMLA LQDFVYISSMDQSSDLDQKRLRNLIVLENAOAVINIOHCAASYDAFSLT CCFVCSPVGFPHEDCSVMKQRPAPQASRSTELLRCQSYAGHFRATRSIADKRPPI TVHEDSISPERQVLIKFSIISPERSASVWGSSCCRRHWHLHSGKNEEVNMGSI " join(37853..38112,38173..38460,38710..38992) /gene="CG11307" /note="CG11307 gene product [alt 2]" /codon_start=1 /protein_id="AAF51753.1" /db_xref="GI:7296467" /db_xref="FLYBASE:FBgn0037106" /translation="MEELCNSSGHSATSSSGNSSISAKTALSCSAATNYLSALNN LCTAGSKLAHSIAVLEQNSISERLENNYTYLJNSWDLARATTVATGVTKTHMLA	
	Alignment Scores: Pred. No.: 7.4e-131 Length: 294308 Score: 2219.50 Matches: 482 Percent Similarity: 53.66% Conservative: 163 Best Local Similarity: 40.10% Mismatches: 271 Query Match: 39.51% Indels: 287 DB: 3 Gaps: 21			US-09-830-837-6 (1-1052) x AE003595 (1-294308) QY 47 PheSerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAla 66 DB 36738 TTTAAACAGCGCTGTGTCCTCAAGTGTCTCACTTCCTCACTCAAAATGCTTGC 36679 QY 67 LysAlaArgAsnSerPheIleSerSerAlaLeuIleSerSerGluVal----- 82 DB 36678 CGGTCGCGAAGTCTACATCGCAAAACTCTTGGTTCAACAGTAAGTTGTATCA 36619 QY 82 ----- 82 DB 36618 AATGGTTATATTATACCTTAAATGCTATCTCTCTGTTTATTTCAACGGATGAGCT 36559 QY 82 ----- 82 DB 36558 GCTTCTTTGAAAGACACAAACGATCTTTCTGACAGAAAGTACCTTTTAAAGTTT 36499 QY 83 ----- Asp-AsnTrpArgI 87 DB 36498 AAACACTTTTCAACAGCTCTTTTATATATCCCTTCCATTTTAGTAAACGAAGTGGAA 36439 QY 87 eileProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLy 107 DB 36438 TGTTCCCGCTCTAAATTTGGCTTGGCAATATCCCAAGTATTGTATATCTTACGAGTTG 36379 QY 107 sGluLysGlnIleysAlaGly-----LeuLeuThrLeuGluAspHisProAsnI 123 DB 36378 CGACGGTTATGATCATCATCATCAGAGTTTATTATAGAAAGGCTTCAGACTCCCATCAGT 36319 QY 123 elysArgValThrProGlnArgIleValPheArgSerLeuIleYtyr---AlaGluSerAs 142 DB 36318 AAAGCGATAGTTCTCCCGCAGCGAGCTAGCAAGGATCTCTAACTACTGACCTATGACAA 36259 QY 142 pProThrValProCysAsnGluThrArgTyrSerGlnIleYtrpGlnSerSerArgProLe 162 DB 36258 CCTAACG----- 36252 QY 162 uArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSe 182 DB 36251 -----TATATTAC-----CGCCATCCCA 36232 QY 182 rArgArgLeuLeuArgAlaIlePro-----ArgGlnValAlaGlnThrLe 197 DB 36231 AGGAGTGCTAGGACAGAACCAACCAACGATCGCCACGACAAATTTGTCGCTACT 36172 QY 197 uGlnAlaaspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPh 217	
mRNA	/codon_start=1 /protein_id="AAF51751.1" /db_xref="GI:7296465" /db_xref="FLYBASE:FBan0007807" /db_xref="FLYBASE:FBgn0023417" /translation="MTLAYVDMPAVIOVQERLSHGSHGLSSGSAAYTHSGGHTGRS GPATGSHGSHSTHGAATMHQSLQSDQFPYPPFPHHSTQSPQOQNHGALLEYLGT RSDYGAISAGAAGHCTRLGHEHESLALHOALQNAVDDVQAPALQNDNVAFMSDPLIKS MKSCKAGNIGSGSPSVFCAVGPLSLISLSTKYKVTIAEVRRLSPPECLNALSILG GVLRAKSNKGRLLREKLEKIGLNLPGARRKAAANTLLTSVVEGATHLAKDFHVC ETEPAROLAIEYVHOTEQSDYSRRKELILHSQOITKELMQLISQDRTTHFGRSOH LLEPSMORHLTHSLTHGRGSPAIMAVLHAFQIFLNSLYLEKLYPSNGGGMVSS LDKSKIDNEKK" complement(<33461..>36792) /gene="CG7169" /map="78B4-78E5" /db_xref="FLYBASE:FBan0007169" /db_xref="FLYBASE:FBgn0037105" complement(join(<33461..34108,34165..34904,35005..35857, 35997..36455,36634..>36792)) /gene="CG7169" /product="CG22149" /db_xref="FLYBASE:FBan0007169" /db_xref="FLYBASE:FBgn0037105" complement(join(<33461..34108,34165..34904,35005..35857, 35997..36455,36634..>36792)) /gene="CG7169" /note="CG7169 gene product" /codon_start=1 /protein_id="AAF51752.2" /db_xref="GI:10726942" /db_xref="FLYBASE:FBan0007169" /db_xref="FLYBASE:FBgn0037105" /translation="MNVFTFLFIISAICSDLAFTAVVPNEFIVHFSKYFAPVRESY IAAKLGSNNWNRIVPRNLNWOYPSDEFILRVCDSGESSEFIERLQTHPSVKAY VQPSVRILNYDAYSLNLYTHRPQGVLRNPNRNDHRQLCSVLHANNILWKLGIIG KGVAIFDTGLTKNHPFRNKRERNTWNEKSLDDRVSHGTGAGVSYTSWFLDAFN YAIRKINILNLSIGGDPFMSPEVEKLELSANNVIMISAAGNDGPLYGLTILNPPGO SDVVGVOFDDKIAKFSRSGMTTWELPGLYGRMLDIVYQSGVSDVRKGCRL LKSMSLLSYKPKITLIPALDFTQNYMPPYSOPLYGSSVAIANVTILNGISVTS HVGIPKWDPENQGVQVSPVIVPWTGMSVFAVKKEGNEGVCVGSIT LVLSFKQTNETHYTEVDPLTKVTPKPNKRILWDOYHSLRYPYPRIPRODLKAV KLPLDMRADHILHTFRDMTHLRNNGYIIDVLRPEPTCFNADSYGALLIVDPERGFG DEELNALQENYKRLGNVVFQWNTYVKKIKFFEDTROWMTDGTGGMANIPALND LLKPGIAGDFVGRGFKLGDHSMYASGATVIFKPMNPGDILIVGTLKNDGILSIIN SKTSPKVALDVPFQFQKANSIOSNEEIVVNAESIAFAIPTDYSTFKNRVILLR TKQBSIFSAKSNHETKNEGLIAYGDSNCLDSTHLKACYLLITFLDFAINSHKSS LLQNLNRIEFHKLERAPLRISQSLIKRSQDNCKQFKWLAPTQKQNAERKSSI IDVILENKGIVLHVLQSLIFVCFRTDRKFNQKFIG" /db_xref="FLYBASE:FBan0011307" /db_xref="FLYBASE:FBgn0037106" join(37155..37175,37839..38112,38173..38460,38710..39358) /gene="CG11307" /product="Ct43349" /db_xref="FLYBASE:FBan0011307" /db_xref="FLYBASE:FBgn0037106" join(37159..37367,37839..38112,38173..38460,38710..39358) /gene="CG11307" /product="Ct31547" /db_xref="FLYBASE:FBan0011307" /db_xref="FLYBASE:FBgn0037106" join(37853..38112,38173..38460,38710..38992) /gene="CG11307" /note="CG11307 gene product [alt 1]" /codon_start=1 /protein_id="AAG22183.1" /db_xref="GI:10726943" /db_xref="FLYBASE:FBan0011307"		mRNA	/db_xref="FLYBASE:FBan0007169" /db_xref="FLYBASE:FBgn0037105" complement(join(<33461..34108,34165..34904,35005..35857, 35997..36455,36634..>36792)) /gene="CG7169" /product="CG22149" /db_xref="FLYBASE:FBan0007169" /db_xref="FLYBASE:FBgn0037105" complement(join(<33461..34108,34165..34904,35005..35857, 35997..36455,36634..>36792)) /gene="CG7169" /note="CG7169 gene product" /codon_start=1 /protein_id="AAF51752.2" /db_xref="GI:10726942" /db_xref="FLYBASE:FBan0007169" /db_xref="FLYBASE:FBgn0037105" /translation="MNVFTFLFIISAICSDLAFTAVVPNEFIVHFSKYFAPVRESY IAAKLGSNNWNRIVPRNLNWOYPSDEFILRVCDSGESSEFIERLQTHPSVKAY VQPSVRILNYDAYSLNLYTHRPQGVLRNPNRNDHRQLCSVLHANNILWKLGIIG KGVAIFDTGLTKNHPFRNKRERNTWNEKSLDDRVSHGTGAGVSYTSWFLDAFN YAIRKINILNLSIGGDPFMSPEVEKLELSANNVIMISAAGNDGPLYGLTILNPPGO SDVVGVOFDDKIAKFSRSGMTTWELPGLYGRMLDIVYQSGVSDVRKGCRL LKSMSLLSYKPKITLIPALDFTQNYMPPYSOPLYGSSVAIANVTILNGISVTS HVGIPKWDPENQGVQVSPVIVPWTGMSVFAVKKEGNEGVCVGSIT LVLSFKQTNETHYTEVDPLTKVTPKPNKRILWDOYHSLRYPYPRIPRODLKAV KLPLDMRADHILHTFRDMTHLRNNGYIIDVLRPEPTCFNADSYGALLIVDPERGFG DEELNALQENYKRLGNVVFQWNTYVKKIKFFEDTROWMTDGTGGMANIPALND LLKPGIAGDFVGRGFKLGDHSMYASGATVIFKPMNPGDILIVGTLKNDGILSIIN SKTSPKVALDVPFQFQKANSIOSNEEIVVNAESIAFAIPTDYSTFKNRVILLR TKQBSIFSAKSNHETKNEGLIAYGDSNCLDSTHLKACYLLITFLDFAINSHKSS LLQNLNRIEFHKLERAPLRISQSLIKRSQDNCKQFKWLAPTQKQNAERKSSI IDVILENKGIVLHVLQSLIFVCFRTDRKFNQKFIG" /db_xref="FLYBASE:FBan0011307" /db_xref="FLYBASE:FBgn0037106" join(37155..37175,37839..38112,38173..38460,38710..39358) /gene="CG11307" /product="Ct43349" /db_xref="FLYBASE:FBan0011307" /db_xref="FLYBASE:FBgn0037106" join(37159..37367,37839..38112,38173..38460,38710..39358) /gene="CG11307" /product="Ct31547" /db_xref="FLYBASE:FBan0011307" /db_xref="FLYBASE:FBgn0037106" join(37853..38112,38173..38460,38710..38992) /gene="CG11307" /note="CG11307 gene product [alt 1]" /codon_start=1 /protein_id="AAG22183.1" /db_xref="GI:10726943" /db_xref="FLYBASE:FBan0011307"	


```
Db 36171 | |||:|||||:|||||:||||| ||||| |||:|||||:||||| 36112
Qy eAspThrGlySerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTr 237
Db 36111 CGACACGGCTACCAACCAACCATCCACACTTCGAAATGTAAGGAAGCAACAACG 36052
Qy pThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValII 257
Db 36051 GACGAATGAAGTCACCTGACGACAGAGTCAGTCATGGCACCCTCGTCGCCGGGTAA 35992
Qy eAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisLysPheArgVa 277
Db 35991 CGCTTCTTCCAGGGAATGCTAGCTTCGCTCCGACGCCGATCTTTACATATTTAAAGT 35932
Qy lPheThrAsn- 280
Db 35931 TTTTACGAAGCTCCCAAGTAGCAAGCTAGAACTGAAATTAACCTAAGCAATCAATTAG 35872
Qy 281 -----Asn-GlnValSerTyrThrSerpPheLeuAspAlaPheAsnTyrAlaIleL 298
Db 35871 CCCTACATCGCGAGGTTCTTACACTTCCTGGTTCCTGGATGCATTCACATACCGATAT 35812
Qy euLysLysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProp 318
Db 35811 ATAGGAAATTAACATCTCAACCTTAGCATTTGGGGTCCCGACTTTATGAGCTCGCGCT 35752
Qy heValAspLysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyA 338
Db 35751 TCCTTGAAAGGTGTGGAACTGTCGCTAATATGTCATATGATTCGCGACGAGCAAG 35692
Qy snAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyV 358
Db 35691 ATGATGTCCTTGTGCGGCACGCTAAACAATCTTGGCGATCAGACGATGATGTGGCG 35632
Qy 358 alGlyGlyIleAspPheGluAsnIleAlaArgPheSerSerArgGlyMetThrThrT 378
Db 35631 TTGGTGGCATTCAGTTTGATGATTAATCCCAAGTTTATGTCGAGAGGAATGACAACGT 35572
Qy rpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyV 398
Db 35571 GGAACCTCCCTTAGGCTACGAGCATGATGGGACTCGATATTTGTCAGTACGGAAGTCAAG 35512
Qy 398 alArgLysGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerP 418
Db 35511 TGAAGGACGTGATGTGCGCAAGGGTGCAGACACTCTCTGGAACATCGTGTCTCTC 35452
Qy 418 roValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGluLeuV 437
Db 35451 CAGTTTGTGAGGGGTGCTGCACCTGCTATTAAGCGGTGCATTTTCAGAAAATCGACTACA 35392
Qy 437 alAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValA 457
Db 35391 TAAACCCAGCATCTCTTAGCAGGTACTCATGGAAGTCCGAGAAACTGCCCATATATA 35332
Qy 457 snMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnS 477
Db 35331 ACATGTTTGAGCAGGGAGCTGGAAACTGAATTTGCTGAAGATGACAGCTATTGCTGT 35272
Qy 477 erTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrM 497
Db 35271 CATACAAACCAAGATAACCTTATTTCGCGCATACCTTGACTTCACCCCAA---AACTATA 35215
Qy 497 etTrpProTyrCysSerGlnProIleTyrTyrGlyMetProThrValValAsnValT 517
Db 35214 TGTGGCTCTTATAGCTCCCAACCTCTGTACTATGAAGCTCCGCGCATATTGCAACGTTA 35155
Qy 517 hrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTyrGlnProT 537
Db 35154 CCATACCTCAATGATCTCTGTGCACAACTCATATAGTTGGCATCCCTCAATTAATGGAATCCCG 35095
Qy 537 yrLeuProGlnAsnGlyAsnIleGluValAlaPheSerTyrSerValLeuTrpP 557
|||:|||||:|||||:||||| |||:|||||:|||||:|||||
```

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: twilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by
M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>= 30); an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION:

The 5' clone is T020D01; 3' clone is F028I16. Actual start of this
clone is at base position 1 of T29J13; actual end is at 100592 of
T29J13.

FEATURES
source
1. 100592
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="V"
/map="unknown"
/clone="T29J13"
BASE COUNT 30991 a 18500 c 18948 g 32153 t
ORIGIN
Alignment Scores:
Pred. No.: 3,82e-115 Length: 100592
Score: 1965.50 Matches: 447
Percent Similarity: 49.09% Conservative: 149
Best Local Similarity: 36.82% Mismatches: 255
Query Match: 34.99% Indels: 368
DB: 8 Gaps: 23

US-09-830-837-6 (1-1052) x T29J13 (1-100592)
QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
DB 35394 TCTGAGATCAAAACGAAGACTAACTATATATCGATTAAAGCAGTATAAGCGCGAAA 35335
QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAlaAspAsnTrpArgIle 87
DB 35334 GATCATCGAATTTACCTCGAATCGAAGCTCGATCCGCTGTTGGGTTGG 35284
QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
DB 35283 ATCGAGAGATTAAACCGCGGACGAAGTATCCGACTGACTCGGTGTTCTATGATCGAG 35224
QY 108 GlnLysGlnLysAlaGlyLeuLeu---ThrLeuGluAspHisProAsnIleLysArgVal 126
DB 35223 GAACCGGGAAGAGCCCGTGGTTGGAGAGATTGAGAGCTGGAGATGGTGAACAGCTG 35164
QY 127 ThrProGln----- 129
DB 35163 AACGTTGAGTTTAAGTACCAGAGAGCTTTTGTCTGCGAGGATCTTTTCCCGACGGGAGAAA 35104

QY 835 ----- 835
DB 33964 GATTACTCCACATTAAGAACCGGTTTGTCTACTCGCAACGAACGAACGATATCACT 33905
QY 836 ----- 836
DB 33904 TTTGCGAAAGCAATAATCATGAACATAAGAACTGAAGGACGTTATGCGGTATATGGGAC 33845
QY 847 SerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeu 866
DB 33844 TCCAACTGCTCGACTCCACGCACTCGAGAGCGTTGCTACTGCTGCTAATAAGCTTT 33785
QY 867 LeuGlnTyrThrSerTyrGlyValThrProSerLeuSerHisSerGlyAsnArgGln 886
DB 33784 TTAGATTTTGCATAAAC-----TCGCAC----- 33761
QY 887 ArgProProSerGlyAlaGlySerValThrProLuarGmetGluGlyAsnHisLeuHis 906
DB 33760 ----- 33760
QY 907 ArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLysProArgProLeuProAla 926
DB 33742 AATCTAAATCGTATACTGAATTCACAAATTAGAGAGACCACTACCCCTTAGGATA 33683
QY 927 ----- 33683
DB 33682 TCGCAAGTATATAAAATCTCGTTTCACAGGACAATAATTTGTGAACAAATTAAGTGGCTT 33623
QY 934 LysProGlnProLeuAsnGlnThr-AlaProSerAsnLeu----- 946
DB 33622 GCACCGGAGCAAAATAACCGCGAGAAAGAAATCTTCTATAATAGACGTAACCATTA 33563
QY 947 TrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysValValLeuProAsnPheArg 966
DB 33562 CTGGAATTAAGAGTATTATATACATTTAGTCAACATCAAACTCATATTTGTTGT 33503
QY 966 gSerAsnArgPro-----GlnValArgProLeuSerProGlyGluSerGlyAlaTrpAs 984
DB 33502 TTCAGACACAGAGATAAATTAATCAAAATTTATTTGGTGGAGGATCGCAAACTAGG 33443
QY 984 pileProGlyLyleMetProGlyArgTyrAsnGlnGluValGlyGlnThrIleProVa 1004
DB 33442 GCAAACAACTATTATTACAGGAGTGCATCCGGATAGTCTAATGACTCCCAATTAT 33383
QY 1004 lPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValGlnIleAsnLys 1023
DB 33382 ATCTTAATTAATTAAGCTAATGTTATCATGTTATTTGTTTATTTTAAAGCGT 33325
RESULT 14
T29J13/c
LOCUS T29J13 100592 bp DNA linear PLN 23-AUG-2000
DEFINITION Arabidopsis thaliana BAC T29J13.
ACCESSION AF296838
VERSION AF296838.1 GI:9885851
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 100592)
AUTHORS Latreil,P., Courtney,L., Abbott,A., Gregory,S. and Bielicki,L.
TITLE The sequence of A. thaliana T29J13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 100592)
AUTHORS Washington University Genome Sequencing Center.
TITLE The A. thaliana Genome Sequencing Project
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 100592)
AUTHORS Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-2000) Department of Genetics, Washington

```
QY 130 -----ArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCys 147
Db 35103 CGTCCGGGGAAGATCTTCTATGCTTTTGAAGAAGAACTCAGTCTCTCTCTATG 35044
QY 148 AsnGluThr-----ArgTrpSerGlnLysTrpGlnSerSerArgProLeu 162
Db 35043 GCGGATACAAAGTAACACCAGGTAAACTGGAGTCGACA-----TCTTCTCGCTCAGGT 34991
QY 163 ArgArg-----AlaSerArgSerLeuGly 170
Db 34990 GAGAAATGCTGAAGTAACATGATCAATATCTGTTTCTCTGCTCACTATCGATCTATGT 34931
QY 171 SerGlyPheTrpHisAlaThrGlyArgHisSerSerArgLeu-----LeuArgAlaIle 189
Db 34930 GCTAATTTTGGTAT-----CGTAAACTGTAAGTGAGGTCTTTA 34892
QY 190 ProArg-----GlnValAlaGlnThrLeuGlnAla 199
Db 34891 AT-AAGTTACATTTGTTTCTATTTGCGAAGACTCAAGTCAAGTCCATGTTTGAGCA 34833
QY 200 AspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThr 219
Db 34832 GATCATCTCTGGAAGAAGGTACACCGCGCTAAAGTTAAAGTGAATATTTGATACT 34773
QY 220 GlyLeuSerGlnLysHisProHisPheLys----- 229
Db 34772 GGAATAAGAGCTGATCACCTCATTTTCG-TAAGATCAAGGTATCATCTTTTGGAGAAAC 34714
QY 229 ----- 229
Db 34713 TGATAAACCATAGTATTTCTCAAGATTTCTGAGAAAGTGTGAAGTGTAAATTTACTAGT 34654
QY 230 -----AsnValLys 232
Db 34653 TTTCCTTTTATCTGCTTGTAGCTCTGAAGTGTGATTCGCGGAGTAACCTTCAG 34594
QY 233 GluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPhe 252
Db 34593 GAGCGTACGAATTTGGACAAATGAGGACACTTTGAATGACAACCTGGTGTATGGAGCTTT 34534
QY 253 ValAlaGlyValIleAlaSerMetArg--GluCysGlnGlyPheAlaProAspAlaGlu 271
Db 34533 GTTCGTGGGTATTTCTGCTGCGAAATCCAGAAATGCTTGGTTTTCCTCCACAGACGGAA 34474
QY 272 LeuHisIlePheArgValPheThrAsn-Asn----- 281
Db 34473 ATCTATGCTTTTCGAGTGTTCACAGATGCCAGGTAATGCATGATCACACAACAGATACT 34414
QY 281 ----- 281
Db 34413 TCTTCTCATCGCGGATAAGGTTTAAACATTCCTTCAACTAAACTAAATCCACTCATGGTCC 34354
QY 281 ----- 281
Db 34353 AGTAAATCTGTTTCTGTTTCCCTAATACACTTGTGTAAGATACTGATATTTGTTACACA 34294
QY 281 ----- 281
Db 34293 GTGATCTTTATTTTGTCCCTCGAGACACTGTAATTTGAGCGCTATAAATCTGTGTGCTTCC 34234
QY 281 ----- 281
Db 34233 TTTTCTCTTTCCAGATACGCAAGTTGTCTTAATCAAAATAGGTTATGATAAGTTTTCGCT 34174
QY 282 -----GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLe 298
Db 34173 CAGCTTGTTCAGGATATACACCTCATGTTTCTTGATGCTTTCAATATGCTATAGC 34114
QY 298 uLysLysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProH 318
Db 34113 AACAGATATGGATGATTAATTTAGCATTTGGGGACCGGATTAATTGGAATCTGCTT 34054
QY 318 eValAspLys----- 321
Db 34053 TGTAGAGAAAGTAATTTGTTTCTTAAATCTCTTGGTTGGAGTTTACTTATCTCTGTAGTCG 33994
QY 321 ----- 321
Db 33993 AGTGTGGGGATGTTTCTCTCTAGCATATATATGAACAGTCTGTGTCTCCAGAGCAGCTGAT 33934
QY 322 -----ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleG 337
Db 33933 TTTCTTCTAACAGTTTGGGAAATAACAGCCAGCATATCATTTATGGTGTCTGCAATGG 33874
QY 337 yAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleG 357
Db 33873 AAATGATGGCCACTTTAGGAACTTAAATAATCCAGCTGACAGAGTGTGTAATAGG 33814
QY 357 yValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTh 377
Db 33813 TGTGGAGGATTTGACAATGATGATCATAGTCTTATCTCATCTCGTGGCATGAGCAC 33754
QY 377 rTrpGluLeuProGly----- 382
Db 33753 CTGGGAACCTTCTCA-TGGGTACTGTATTTCTTGTGCTTCTACATTCCTCTTTACAAA 33695
QY 382 ----- 382
Db 33694 CCCTACAATCGTAAAGATACAAAATAGTTTCTTGTGCAATTCCTAACTTCAAAGTTT 33635
QY 383 -----GlyTyrGlyArgMetLysProaspIleValThrTyrGlyAlaGlyValArgG 400
Db 33634 TGTTCACAGATATGGCGGTGAACACCATGTTGTGTCATATGGCCGCCACATATGGG 33575
QY 400 ySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVa 420
Db 33574 GTCCAAGATCAGCAGCTGGTTGTAAAGCCTGTCTGGAACAAGTGTGCTAGTCTGTGT 33515
QY 420 lAlaGlyAlaValThrLeuLeuValSerThrVal-----GlnLysArgGluLeuVa 437
Db 33514 TGTGTGTTATTTGTTGCTACTTCTTAAGTGTATTCTCTGAAGCTAGAGAAGATCTGCT 33455
QY 437 lAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValas 457
Db 33454 CAATCAGCAAGCATGAAGCAGCATGGTGCAGAGGTGCTGCTGCTGCTGCTGCTGCTGCT 33395
QY 457 nMetPheGluGlnGlyHisGlyLysLeuAspLeu----- 468
Db 33394 TATGTATGAACAGGCTGCTGGAAGAGTGTGATCT-GTAAAGTTACTTATCCATATTTGTTAT 33336
QY 468 ----- 468
Db 33335 CACCATTGTACTCTTTATTTTATTTTAACTTAGTCTCTTATTTTGTAGTCTCTTAAT 33276
QY 469 -----LeuArgAlaTyrGlnIleLeuAsnSerTyrLysProGlnAl 482
Db 33275 TTTTCCACTCTTTCTCAGATTAATCATATGAATCTAAAGAGCTACACCCCGGCG 33216
QY 482 aSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCys 502
Db 33215 AAGCAATTTCCCGAGCATTTCTTGATTAATATGATGTGCTCATATCTCTGGCCCTTTTGTG 33156
QY 502 rGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsnGlyme 522
Db 33155 TCAGCCGCTATATGCGCGGCTATGCCCATCATTTTCAACACGACAAATCTTAAATGATAT 33096
QY 522 tGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGlnAsnG 542
Db 33095 GGTGTCTATGGCTATATTTGAAGTCCACCAACATGGCACCT---GCAACAGGAGGAGG 33039
QY 542 yAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGlyTyrLe 562
Db 33038 AAATCTTCTGAGCATTCACCTTTAAGTACCCGGATGTTATATGATGCGCTGGAGTGTATCT 32979
QY 562 uAlaIleSerIleSerValThrLysLysAlaLeuSerTrpGluGlyIleAlaGlnGlyH 582
Db 582 uAlaIleSerIleSerValThrLysLysAlaLeuSerTrpGluGlyIleAlaGlnGlyH 582
```

```
Db 32978 AGCTTACATCAGATCAAGGAAGAGTGCACAGTTTCACAGGTGAATAGAGGTAA 32919
Qy 582 sValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluInTh 602
Db 32918 TGTAACTGTAAAGTCTATAGCCCA-----CCAGCCTCTGGAGAAAGTGGGCTAGAG 32865
Qy 602 rSerThrValLysLeuProIleLysValLysIleLeuProThrProArgSerLysAr 622
Db 32864 GAGTACTTGTCTCTCAGTTGAAACTGAAAGTGATTCCTCACTCCACAGCGGAAAGC 32805
Qy 622 gValLeuTrpAspGlnThrHisAsnLeuArgTyProProGlyTyThrPheProArgAspAs 642
Db 32804 TATTTTGGGATCAATTTACAGACATTAATATCTCTCTGTTATATTCAGAGACTC 32745
Qy 642 nLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheAr 662
Db 32744 TTTGATCTCGCANTGACATATCTTGTGCAAGTGTACCTGCTGATCAACATTTCA 32685
Qy 662 gAspMetTyrglnHisLeuArgSerMetGlyTyThrPheValGluValLeuGlyAlaProPh 682
Db 32684 CATCATGTACACATCTTACGTGATGTGGGTACTATATCGAGACTCTTGGTTGCGCCCT 32625
Qy 682 eThrCysPheAspAlaSerGlnTyrglyThrLeuLeuMetValAspSerGluGluGluTy 702
Db 32624 TACATGTTTGTATGCTCAGCAATATGAACCTTTTGTGATGTGACCTTGAAGATGATTA 32565
Qy 702 rPheProGluGluIleAlaLysLeuArgAspValAspAsn---GlyLeuSerLeuVa 721
Db 32564 CTTTCCAGAGAATTTGAAAAACTCAGAGATGACGTTATCATATCTGACTGGCTGGT 32505
Qy 721 lIlePheSerAspTrpTyrrAsnThrSerValMetArgLysValLysPheTyrrAspGluAs 741
Db 32504 TGTGTTCTGAGTGTATATGTGATACCATGTTGAGATGAGATTTCTGATGATAA 32445
Qy 741 nThrArgGlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLe 761
Db 32444 CACACGAGTTGTGGACTCCAGTCACTGGAGTGCATATATCCAGCACTTAATAACT 32385
Qy 761 uLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrglyGluPheThrLeuAl 781
Db 32384 TCTTCATCTTTGGGATGCGTTTGGGGCAAAATCTGATGGGATTTCAATTTGA 32325
Qy 781 aAsnHisAspMetTyrrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyVa 801
Db 32324 CGTGAGCAGAGTCGATATGCTTCGGACAAACATGTCGGTTTCTCTGCTGGTGGTT 32265
Qy 801 lValIleThrGlnThrPheLysAsp----- 809
Db 32264 TTTGCACACTTTTCCCTTTATTAGACAGCTCTGAGATGGCGCTACCCAAATCTACTGCT 32205
Qy 810 ----GlnGlyLeuGluValLeuLysGlnGluThrAlaValValGluAsnValPro----- 826
Db 32204 AACTGAGCATCAAAAGGTAATGAGAAACCAGTTCTCTCCACTTCTAACTGTGCTTGGAT 32145
Qy 826 ----- 826
Db 32144 TTTCTTAATTAATTCAGTGATTTGTTGTGATATGTTGGCTGTGTTTACAGGAAGACC 32085
Qy 827 ----lLeuGlyLeuTyrglnIleProAlaGluGlyGlyValGlyValLeuTyrg 845
Db 32084 CTGCTGTTCTTGGCTTTTGGAGATT-----CGTCAAGGTCTGGTGGTGTATATG 32034
Qy 845 lYAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTrpLeuAspA 865
Db 32034 GAGACTCAAAATTTTGGATAGTACCACATGTCACACTGTCACACTGCTCTCTGAAGA 31974
Qy 865 lAlaLeuGlnTyrrThrSerTyrglyValThrProProSerLeu---SerHisSerGlyA 884
Db 31974 AAATCTAGATTTTACGATCTCAATATCAAGACCCCTGTACTTTTCTCAAAATTTGCTA 31914
Qy 884 snArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMetGluGlyAsnH 904
Db 31913 AGAGATATTCACCAAGTAATCATAGACAGGAAGCAACATGCCATCCCGAAGCAATGATGTA 31854
```

```
Qy 904 isLeuHisArgTyrrSerLysValLeuGluAlaHisLeu 916
Db 31853 ATTTTTCACATATCTTCTGTGATCGGAAAGGAGCTA 31816

RESULT 15
LOCUS AC017581 115884 bp DNA linear HTG 10-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
AC017581
ACCESSION AC017581
VERSION AC017581.1 GI:6554416
KEYWORDS HTG: HTGS_PHASE2
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyroidea; Drosophilidae; Drosophila.
Adams,M. and Venter,J.C.
REFERENCE Direct Submission
AUTHORS Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
TITLE Rockville, MD, USA
JOURNAL This sequence was identified as CDM:10211360 by the submitter.
COMMENT For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Location/Qualifiers
source 1. 115884
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 33854 a 24803 c 24145 g 33082 t
ORIGIN
Alignment Scores:
Pred No.: 1 35e-102 Length: 115884
Score: 1772.00 Matches: 375
Percent Similarity: 56.61% Conservative: 122
Best Local Similarity: 42.71% Mismatches: 216
Query Match: 31.55% Indels: 167
Db: 2 Gaps: 14

US-09-830-837-6 (1-1052) x AC017581 (1-115884)
Qy 280 AsnAsnGlnValSerTyrrThrSerTrpPheLeuAspAlaPheAsnTyrrAlaIleLeuLys 299
Db 7 CATCGCAGGTTTCTTACACTTCTCGTTCCTGGATGCATCACTACGATATATAGG 66
Qy 300 LysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProPheVal 319
Db 67 AAAATAACATTTCTCAACCTTAGCATTTGGGGTCCCGACTTATGGACTCGCCCTTGGTT 126
Qy 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
Db 127 GAAAGGTGTGGAACTGTGCGCTAATAATGTCATATATGATATCGCAGCAGGAATGAT 186
Qy 340 GlyProLeuTyrglyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
Db 187 GTCCTTGTACGGCAGCTTAACAATCTCTGGCAGCAGCATGATAGTTGGCTGGT 246
Qy 360 GlyIleAspPheGlnAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu 379
Db 247 GCATTCAGTTGATTAATTCGCAAGTTGATTCGAGAGGATGACAACTGGGAA 306
Qy 380 LeuProGlyTyrglyArgMetLysProAspIleValThrGlyAlaGlyValArg 399
Db 307 CTTCCCTTAGCTACGAGCTAAGGACTCGATATTTGTCAGTCGAGGATCAAGTGGAA 366
Qy 400 GlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProVal 419
Db 367 GGCAGTGTATGCGCAAGGGTGCAGACGACTCTCTGGAACATCCGCTGCTCTCCAGTT 426
```

```
QY 420 ValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGluLeuValAsn 438
|||||
Db 427 GTTGAGGGGCTCTCACTGCTTATAGCGGTGCATTCAGAAATCGACTACATAAAC 486
|||||
QY 439 ProAlaSerMetLysGlnAlaLeuLeuAlaSerAlaArgLeuProGlyValAsnMet 458
|||||
Db 487 CCAGCATCTCTTAAGCAGGTACTCATTTGAAGGTGCGGAGAACTGCGCATTTAATCAG 546
|||||
QY 459 PheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyr 478
|||||
Db 547 TTGACAGGAGCTGGAATACTGAATTTGCTGAAGAGATGATGCTATTTGCTGCATAC 606
|||||
QY 479 LysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrp 498
|||||
Db 607 AAACCAAGATACACCTTATTCGGCATACCTTGACTTCACCAA---AACTATATGTGG 663
|||||
QY 499 ProTyrCysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIle 518
|||||
Db 664 CCTATAGCTCCCAACCTCTGTAATGTAAGAGCTCGGTCTATTGCAAACTGTACCATTA 723
|||||
QY 519 LeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeu 538
|||||
Db 724 CTCATGGTATCTCTGCACAGTCATATAGTTGGCATCCCTAAATGGATTCGGATTTC 783
|||||
QY 539 ProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrp 558
|||||
Db 784 GAAACCAAGGTCAAGTCTTCAAGTATCTGCACAAGTTTCGCCTATCGTTTGGCGGTG 843
|||||
QY 559 SerGlyTyrLeuAlaIleSerIle----- 566
|||||
Db 844 ACCGGTGGATGTCAGTCTTTTATTTGTTGAGTATGAAATAGCAACGCTATATAAAGGAT 903
|||||
QY 566 ----- 566
|||||
Db 904 ATATGGGTACATATTGATATGAAAGTGAAGCTGAACCACTTCTATACCAACAATTTTGT 963
|||||
QY 567 ---Ser-ValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisValMetI 585
|||||
Db 964 ACCAGCTGTAATAAGAGAGGAGAAACTTTGAAGGTGTTGTAAGAGAGTATCACCT 1023
|||||
QY 585 eThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThrSerThrVa 605
|||||
Db 1024 AGTTTGGAAAGCTTTAAACAGACACCAACCAAGAACT-----CATGTTACAGAAGT 1074
|||||
QY 605 LysLeuProIleLysValLysIleIleProThrProArgSerLysArgValLeuTr 625
|||||
Db 1075 CGACTTCTCTTAACAAATAAGGTACTCCAAACCCGCAAGAAACAGAGGATTTATG 1134
|||||
QY 625 pAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArgAspAsnLeuArgMe 645
|||||
Db 1135 GGATCAGTACCACAGCTAAGGTATCCACCGGCTATATCCACGAGATGATCTCAAAGT 1194
|||||
QY 645 tLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgAspMetTy 665
|||||
Db 1195 TAAACTAGATCTCTCGACTGGAGGCGACAGACCATATACACAAACTTTAGGACATGTA 1254
|||||
QY 665 rGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThrCysph 685
|||||
Db 1255 TACACATTTACGAATTTGGCTGCTACATTTGATTTTGGAGAACCCCTTCACCTGCTT 1314
|||||
QY 685 eAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluTyrPheProGl 705
|||||
Db 1315 CAATGCTCGGATATGGCGCTTATGATTTGACCTTGAGAGGAGGTTTGGCGACGA 1374
|||||
QY 705 uGluIleAlaLysLeuArgArgAspVal---AspAsnGlyLeuSerLeuValIlePheSe 724
|||||
Db 1375 GGAATAAAGCGCTTTACGAAACCGTGATATAAAGAGGCTTGAATGCTGCTGATTCGG 1434
|||||
QY 724 rAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThrArgGl 744
|||||
Db 1435 AGACTGGTATAACACCCTGTGTGATGAAATAATTAATTTCTTTGACGAGAACCCGACA 1494
|||||
```

Db	2495	TAATTATATTAAAGCCTAATGTTATCATGTTATTTGTTTATTTTAAAC	2544
Db	2495	TAATTATATTAAAGCCTAATGTTATCATGTTATTTGTTTATTTTAAAC	2544
RESULT 16			
LOCUS	1650 bp	linear	PRI 18-FEB-2000
DEFINITION	Homo sapiens mRNA; CDNA DKFZp434A219 (from clone DKFZp434A219); partial cds.		
ACCESSION	AL133583		
VERSION	AL133583.1	GI:6599167	
KEYWORDS	Homo sapiens.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
REFERENCE	Poustka, A., Klein, M., Mewes, H.W., Gassenhuber, J. and Wiemann, S.		
AUTHORS	Direct Submission		
TITLE	Submitted (15-DEC-1999) MIPS, Am Klopferspitz 18a, D-82152		
JOURNAL	Martinsried, GERMANY		
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.		
COMMENT	This clone (DKFZp434A219) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.		
FEATURES	Location/Qualifiers		
source	1..1650		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/map="361.5 CR from top of Chr16 linkage group"		
	/clone="DKFZp434A219"		
	/tissue_type="testis"		
	/clone_lib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI"		
	/dev_stage="adult"		
gene	1..957		
CDS	/gene="DKFZp434A219"		
	/note="K1A0091, partly"		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="CA63727.1"		
	/db_xref="GI:6599168"		
	/translation="VKFYDENTRWMMPTGGANIPALNELLSSWNMFSDGLVEGEPTLANHMYASGCSIAKEPDEVVITOTFKDQLEVLKQETA VENVPIGLYQIPAE GGGRIYVLDGNSCLDDSHRQDCFWLLDALQYTSYVTPSLSHSGNRQPPSGAGS VTPRMENHLHRYSKVLEHLGDPKRPPLPACPRLSWAKPOPLNETAPESNLWKHKL LSLDLKVLPLNFRSNRPQVRPLSPGESGANDIPGIMPGRYNOEYVGQIPVFAFLGA MVVLAFFVQINKAKSRPRKRPVRLPQMQVHPPTFSV"		
	1612..1617		
	polyA_signal		
	1640		
	polyA_site		
BASE COUNT	407 a	394 c	448 g
ORIGIN	401 t		
Alignment Scores:			
Pred. No.:	1,496-101	Length:	1650
Score:	1716.00	Matches:	318
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	30.55%	Indels:	0
DB:	9	Gaps:	0
US-09-830-837-6 (1-1052) x HSM801435 (1-1650)			
Qy	735	ValLysPheThrGluAsnThrArgGlnTrpMetProAspThrGlyAlaAsn	754
Db	1	GTGAAGTTTATGATGAACACACAGGCGATGGTGGATCCGGATACCGGAGGAGCTAAC	60
Qy	755	IleProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyr	774
Db	61	ATCCAGCTCTGAATGAGCTGCTGTGTGTGAACATGGGGTTCACGATGGCCGTAT	120
Qy	775	GluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAla	794
Db	121	GAAGGGAGTTCACCTCGCCCAACCATGATGATATATGCGTCAGGTGCACATCGCG	180
Qy	795	LysPheProGluAspGlyValIleThrGlnThrPheLysAspGlnGlyLeuGluVal	814
Db	181	AAAGTTTCCAGAGATGGCGTGTGATACACAGACTTTCAAGGACCAAGGATGGAGGT	240
Qy	815	LeuLysGlnGluThrAlaValGluAsnValProIleLeuGlyLeuTyrGlnIlePro	834
Db	241	TTAAGCAGCAACACAGCTGTTGAAAACGTCCTCCATTTTGGGACTTTATCAGATCCA	300
Qy	835	AlaGluGlyGlyValArgTleValLeuTyrGlyAspSerAsnGlyLeuAspSerHis	854
Db	301	GCTGAGGCTGGAGCGCGATGTTACTGTATGGGACTCCCAATTTGATGACAGTCAC	360
Qy	855	ArgGlnLysAspCysPheThrPheLeuAspAlaLeuGlnTyrThrSerTyrGlyVal	874
Db	361	CAACAGAGGACTGCTTTTGGCTTCTGGATGCTCTCTCCAGTACACATCGTATGGGTG	420
Qy	875	ThrProSerLeuSerHisSerGlyAsnArgGlnArgProSerGlyAlaGlySer	894
Db	421	ACACCGGCTAGCCTCAGTCTGGAACCGCCGCCCTCCCGTGGAGAGGCTCA	480
Qy	895	ValThrProGluArgMetGluGlyAsnHisLeuArgTyrSerLysValLeuGluAla	914
Db	481	GTCACTCCAGAGAGATGGAAGAAACATCTTCATCGGTACTCAAGATTCTGGAGGCC	540
Qy	915	HisLeuGlyAspProLysProArgProLeuProAlaCysProArgLeuSerTrpAlaLys	934
Db	541	CATTTTGGAGACCCAAACCTCGGCTCTACACGCTGCCAGCTTCTCTGGGCCAAG	600
Qy	935	ProGlnProLeuAsnGluThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSer	954
Db	601	CCACAGGCTTTAAACGAGAGCGCGCCCACTAGTAACCTTTGAAACATCAAGAACTACTCTCC	660
Qy	955	IleAspLeuAspLysValValLeuProAsnPheArgSerAsnArgProGlnValArgPro	974
Db	661	ATTGACCTGGACAAGGTGGTGTATACCAACTTTCGATCGAATCGCCCTCAAGTGGAGGCC	720
Qy	975	LeuSerProGlyGluSerGlyAlaTrpAspIleProGlyLysIleMetProGlyArgTyr	994
Db	721	TTGTCCCTGGAGAGCGCGCTGGGACATTCCTGGAGGATCATGCTGGCGGTAC	780
Qy	995	AsnGlnValGlyGlnThrIleProValPheAlaPheLeuGlyAlaMetValValLeu	1014
Db	781	AACAGAGAGTGGGCCAGACCATTCCTGCTTTGCTTCTCTGGAGCCATGGTGTCTGT	840
Qy	1015	AlaPhePheValValGlnIleAsnLysAlaLysSerArgProLysArgAlaGlyProArg	1034
Db	841	GCCTTCTTTGTGGTACAAATCAACAGCCGAGAGAGCCGAGCCGAGGAGGAGGCCAGG	900
Qy	1035	ValLysArgProGlnLeuMetGlnGlnValHisProProLysThrProSerVal	1052
Db	901	GTGAAGCGCCGCGAGCTCATGACGAGGTTTCAACCCGCAAGACCTTCCTGGGTG	954
RESULT 17			
LOCUS	170371 bp	DNA	linear
DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6, clone: P0681F10.		
ACCESSION	AB026295		
VERSION	AB026295.2	GI:5295936	
KEYWORDS	Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA, clone: P0681F10.		
SOURCE	Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA, clone: P0681F10.		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriartoideae; Oryzaceae; Oryza.

1
Sasaki, T., Nagamura, Y. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
clone: P0681F10
Published Only in Database (1999)
2 (bases 1 to 170371)
Sasaki, T., Nagamura, Y. and Yamamoto, K.
Direct Submission
Submitted (21-APR-1999) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Jun 30, 1999 this sequence version replaced gi:4689084.
Detailed information of this PAC DNA sequence is available at
http://www.dna.affrc.go.jp:82/
The orientation of the sequence is from T7 to SP6 of the PAC clone.
Genes were predicted from the integrated results of the
following: GENSCAN1.0, BLASTN1.4, BLASTX2.0 as well as
SplicePredictor (October 1998 version). The genomic sequence was
searched against the non-redundant database NRP (PIR, SWISSPROT,
GENPEPT, PDB) from MAF DNA Bank and the cDNA sequence database at
RGP. Protein similarities of the coding regions were searched
against NRP with BLASTP2.0. ESTs represent the identified cDNA
sequences using BLASTN1.4 with the corresponding DDBJ accession no.
and RGP clone ID.
Sequence updated (17-Jun-1999).
Location/Qualifiers
1. 170371
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="6"
/clone="P0681F10"
complement(join(1731..2330,2700..2948))
/note="hypothetical protein"
/codon_start=1
/db_xref="GI:5295937"
/protein_id="BAA81838.1"
/translation="MAHLSISLFFSDSNKVLPGARRSTSLVGLDRPREARWG
FDGDLHLAASAGNAGDGTAVRAPNGCCVAGHSASAIIGGIVDCTTGEQADF
FYEPGCAKDGLQMDSTLMAATMFLRTDTSFIILDFDRKNNAWLMSDVG
PASLEDKYLHAIAMEDGRCARPLSPAAPRAATLRARHRVGRSLSPQLRRAPCSA
AGRLTOVCHHACSREERKKRDKVGTPTPRRLAVALAEPKLGPAVAHSLCSAARK
KKGKRP"
join(4517..4711,5043..5301,7197..7222)
/note="EST C73147(E2988) corresponds to a region of the
predicted gene.
Similar to Glycin max gmsti mRNA.(X79770)"
/codon_start=1
/protein_id="BAA81839.1"
/db_xref="GI:5295938"
/translation="MSVPRNPWRAAEVQVKRGDSCYKNGSGVKAIKHYTRAE
LDPSDFLIKRAKALSGYQECVDCNDALRRGEELGSGSGNKLISEALLWKAS
ALEHLADCAADYEQVILLRLSLTCHSEEAQIRLKGALFMREQYEELKSHGNSLID"
complement(8389..8799)
/note="ESTs C74053(E30274), D15300(C0425), AU030134(E50667)
correspond to a region of the predicted gene.
Similar to Asparagus officinalis mRNA for Histone
H3.(X82414)"
/codon_start=1
/protein_id="BAA81840.1"
/db_xref="GI:5295939"
/translation="MARTQTARKSTGGKAPRKQLATKAARKSAPATGGVKPHRFRP
GTVALREIKYOKSTELLRKLPQRLVREIAQDFKTLRFOSSAVAALQEAAYLV
GLFEDFNLCIAHAKRTYIMPDKIQLARRINGERA"
9348..9758
/note="ESTs C74053(E30274), D15300(C0425), AU030134(E50667)
correspond to a region of the predicted gene.
Similar to Asparagus officinalis mRNA for Histone
H3.(X82414)"

/codon_start=1
/protein_id="BAA81841.1"
/db_xref="GI:5295940"
/translation="MARTQTARKSTGGKAPRKQLATKAARKSAPATGGVKPHRFRP
GTVALREIKYOKSTELLRKLPQRLVREIAQDFKTLRFOSSAVAALQEAAYLV
GLFEDFNLCIAHAKRTYIMPDKIQLARRINGERA"
complement(join(10487..10755,10866..11136,11240..11391,
11604..11816,11910..12144))
/note="ESTs AU075322(C11109), D22430(C11109) correspond to
a region of the predicted gene.
Similar to Medicago sativa early nodule-specific protein
(ENOD8) gene, complete cds.(U18899)"
/codon_start=1
/protein_id="BAA81842.1"
/db_xref="GI:5295941"
/translation="MRRITWLATAAALCCLVAASATACQRFPAVNFQDSNSDTGG
FWAEPQAQAPFGMTYFRPAGRASPGRVLDVQLQAMGLPILLSPVLOSQSGYRGA
NPAFLASTALQPNSTLFTVGISPFVFLAVOLNOMKELTKVLSNGNNDQLPADVLIHN
ALYITDIGNDLTSLNGSOSIETVKOSLPSVSKISSTVOELINIGARNIMFNMAPI
GCYPAFLTKLPHSTNDMDGYCMKTYNSAVTYNELLNLSLAERVKKLQADASTVLDK
HATVLELFRPKAHGLKYGTCACCGYGDGAYNFNDVYCGSKLLNGQVTVAKACADP
ONVSWDGIHATEAANKIIASSLSMGSGSYPPFDLSKLCHLOPIA"
complement(join(13415..13545,14288..14479,15426..15472,
16418..16640))
/note="Similar to Gossypium barbadense Sea Island
proline-rich cell wall protein gene complete cds.(U04267)"
/codon_start=1
/protein_id="BAA81843.1"
/db_xref="GI:5295942"
/translation="MDYIRGRPGSGYGGVGGGSSSLIPRVQGPQSHGVANAPPP
EPRAAPYHHGPPTVSAAPHVPVPSSTSMGIQVVIKPAYRITPPQPPQTEIPR
SIFNDFEYERKILAEAKENPNWSKFVIESQPPPPPPGPKLTPTTTSVYKFEVK
SYNALHEMGFTSSNPPELLAIHNDPDKVIOHLIGTS"
complement(17973..18149)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA81844.1"
/db_xref="GI:5295943"
/translation="MPIEPRAGNPEVAPRVWPDDPCRSALQPTQRTKTPAEMGI
EGHRSSAVATPPDT"
join(20039..20155,20868..21914)
/note="Similar to Nicotiana tabacum mRNA for ERP1,
complete cds.(D38123)"
/codon_start=1
/protein_id="BAA81845.1"
/db_xref="GI:5295944"
/translation="WYNRASETAHTVSAHTSSLSSPMVEIVVEKKARRIVRGW
VRSNAAAAAFAVAPAPRVVRLCDHDDATDSDGDDGDDAPRARLLVHEIHVA
RQPVAMSAASQRRRQPMKRTESAVDAMDATAPAAERKFRGVRKRPWKYGAIE
IRVSOQSRVWLTFTAEAAVYDHAALRLRGPSTTFNPMTPAAPSPPPSRATYA
GAAGYDESDESQVLGSPVSLRPMPARATAKKEAEDDSAPDILGISAGDGLISP
FTCDVLPFPDDEMGFGSFGPEPTPPMVFDDDCMARLGHAPNDDEHPVTSSFLD
GDLGLSPSTEVDGFFSDYGGDDLFAAEPEPAL"
join(24131..24185,25180..25235,25645..25665)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA81846.1"
/db_xref="GI:5295945"
/translation="MAATEAVTTDDATMVVCMATMAATEVVTDDMTMVGSAVR"
complement(join(27565..28664,30652..30958))
/note="EST C74734(E50712) corresponds to a region of the
predicted gene.
hypothetical protein"
/codon_start=1
/protein_id="BAA81847.1"
/db_xref="GI:5295946"
/translation="MAAWTAAGMLGSAAYTRMRAENPSEMPRRVILLSHAGLENEID
RGNFVWKAALVYESTVGBVDDHTRLSRLLLNLAAAYCCPHVDPSPRRPQALSE
DEDKKAEDDQDLAKRTOHPDEQTEENKQVIDAFLVVGFLPRLTKAAATGGGKGA
ARRDAVDSFKATHMODLVTDVILKENOLPIKHLAVADLAEAAVHAAGDIPGLKD
DYAKALREYKLGAFARANFGVIRSFCSYSPFSKDEQAKKPDDAISELTLDCILH
ASLVPPSSGAGGAGGVKGYTSRIPTAKLRKRGVRLGAGVEDGRAVQVQFKEDAATL
RLPALVDFDKLATVARNLLARELEEQSKPVTRYFYQLMNLVEEVADVRLRRAGVVRG

GSRGAGEVHELKIKIDGYATYPSPFMAMDVOQVKVFHEKRMNFFVRYRPAIVAAS
SVVAASVAIVATKKRG"

CDS
complement(join(33024. .33137,33976. .34006,34142. .34194,
34276. .34384,34509. .34633,35297. .35386,35465. .35527,
35638. .35729,35826. .35907,36098. .36271,36368. .36449,
37285. .37409,37497. .37574,38232. .38587,39496. .39793))
/notes="ESTs AU075322(C11109),D22430(C11109) correspond to
a region of the predicted gene.

Rice gene for soluble starch synthase (SSS1), complete cds
(exon1-15).(D38221)"

/codon_start=1
/protein_id="BAA81848.1"
/db_xref="GI:5295947"
/translation="MATAAGMGIGACILVAPQVRGRRLRLOVRPRCVAEISDDGGS
AORLPAPLVKQVPLTEFLPTSTPAPTQSPAPPTPPPLDSCVGEIEPDLLEGLI
EDSDIKTIEVASEQSEIMDKVQAOAKVTSVWFVIGEASPYAKSGGLGDDVCSLPI
ALALGRHVWVMPRYMNGALNKNFANFYTEKHKIPCFGGEHEVITFFHEIRUSVDM
VFDHSYRPNLYGDNFGNQFRITLLCYAACAEPLILLEGYYIGOKCMFV
NMHASLVPVLLAKIRPIGVYRDARSVLYVHNLAHQGVPEASTYDGLGPPENYAL
EMFPEWARHLDKGEAVNFKGAVVYADRIYVTSQVYSWEVTTAEGGQGLNELLS
KSVLNGIVNGIDINDMNSTDKPLPYHYSVDLSGKAKKELQKELGLPRDPVPL
IGFRLDYKQKIDILIKLAPLMDRNIQFVMLSGDGPFGFEGMRSESGYRDKFRGM
VGFSPVSHRITAGCDIILMPSPFCPCGNQLYAMQYGVVPHVGGTGLRDTVENFP
FAKGGGCTGWAFSPLETKGIADGNFDIQGTQVLLGSNEARHVKRLYMGPCRLTV"
join(45936. .46041,46221. .46516,46748. .46965,47016. .47577,
48123. .48381,50137. .50150)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA81849.1"
/db_xref="GI:5295948"
/translation="MVVRGGGGAERCGGECRAPAALRRFAGAGPEHGDPLQTEAE
AEGSAVVVGAAGAEVTEGSIQAGATEAGLRNSAVETVAFHRAHGLNRL
RFSRVVVAASLLPLFSLPSVLDVFLQTNRRTPDRHSLSGVMTSLSTRPH
LVSGTITPLPSGGDIVAQDEAPVSEQEARAPFKVLAALAKALPKPNRPSHRAPYS
ASATSAARSPPEAMQFHDHRHVRILSRVHGYLHADEGRVSLQPTGASLTAVM
TVHLEGSPQRLRLHNAAYGRYLAATKPGPSGLGRHVALINDRLDDESISWEAV
RTAKGDDVLLRHATGRNLRAHAGATVDDRYRMLLWQVQVVEAIPSAADSVPRPPI
SRQDGPETIRTFVPAAPNGTIAEDWTFQFGRSVQRLEKETSRAGETQVCVA
GRYGLIPLTHLPHTETIDIIIMWTGTDGR"

CDS

complement(join(50758. .51068,53097. .53115))
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA81850.1"
/db_xref="GI:5295949"
/translation="MATAAGMGIGACILVAPQVRGRRLRLOVRPRCVAEISDDGGS
AORLPAPLVKQVPLTEFLPTSTPAPTQSPAPPTPPPLDSCVGEIEPDLLEGLI
EDSDIKTIEVASEQSEIMDKVQAOAKVTSVWFVIGEASPYAKSGGLGDDVCSLPI
ALALGRHVWVMPRYMNGALNKNFANFYTEKHKIPCFGGEHEVITFFHEIRUSVDM
VFDHSYRPNLYGDNFGNQFRITLLCYAACAEPLILLEGYYIGOKCMFV
NMHASLVPVLLAKIRPIGVYRDARSVLYVHNLAHQGVPEASTYDGLGPPENYAL
EMFPEWARHLDKGEAVNFKGAVVYADRIYVTSQVYSWEVTTAEGGQGLNELLS
KSVLNGIVNGIDINDMNSTDKPLPYHYSVDLSGKAKKELQKELGLPRDPVPL
IGFRLDYKQKIDILIKLAPLMDRNIQFVMLSGDGPFGFEGMRSESGYRDKFRGM
VGFSPVSHRITAGCDIILMPSPFCPCGNQLYAMQYGVVPHVGGTGLRDTVENFP
FAKGGGCTGWAFSPLETKGIADGNFDIQGTQVLLGSNEARHVKRLYMGPCRLTV"
join(45936. .46041,46221. .46516,46748. .46965,47016. .47577,
48123. .48381,50137. .50150)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA81848.1"
/db_xref="GI:5295947"
/translation="MATAAGMGIGACILVAPQVRGRRLRLOVRPRCVAEISDDGGS
AORLPAPLVKQVPLTEFLPTSTPAPTQSPAPPTPPPLDSCVGEIEPDLLEGLI
EDSDIKTIEVASEQSEIMDKVQAOAKVTSVWFVIGEASPYAKSGGLGDDVCSLPI
ALALGRHVWVMPRYMNGALNKNFANFYTEKHKIPCFGGEHEVITFFHEIRUSVDM
VFDHSYRPNLYGDNFGNQFRITLLCYAACAEPLILLEGYYIGOKCMFV
NMHASLVPVLLAKIRPIGVYRDARSVLYVHNLAHQGVPEASTYDGLGPPENYAL
EMFPEWARHLDKGEAVNFKGAVVYADRIYVTSQVYSWEVTTAEGGQGLNELLS
KSVLNGIVNGIDINDMNSTDKPLPYHYSVDLSGKAKKELQKELGLPRDPVPL
IGFRLDYKQKIDILIKLAPLMDRNIQFVMLSGDGPFGFEGMRSESGYRDKFRGM
VGFSPVSHRITAGCDIILMPSPFCPCGNQLYAMQYGVVPHVGGTGLRDTVENFP
FAKGGGCTGWAFSPLETKGIADGNFDIQGTQVLLGSNEARHVKRLYMGPCRLTV"
join(45936. .46041,46221. .46516,46748. .46965,47016. .47577,
48123. .48381,50137. .50150)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA81849.1"
/db_xref="GI:5295948"
/translation="MATAAGMGIGACILVAPQVRGRRLRLOVRPRCVAEISDDGGS
AORLPAPLVKQVPLTEFLPTSTPAPTQSPAPPTPPPLDSCVGEIEPDLLEGLI
EDSDIKTIEVASEQSEIMDKVQAOAKVTSVWFVIGEASPYAKSGGLGDDVCSLPI
ALALGRHVWVMPRYMNGALNKNFANFYTEKHKIPCFGGEHEVITFFHEIRUSVDM
VFDHSYRPNLYGDNFGNQFRITLLCYAACAEPLILLEGYYIGOKCMFV
NMHASLVPVLLAKIRPIGVYRDARSVLYVHNLAHQGVPEASTYDGLGPPENYAL
EMFPEWARHLDKGEAVNFKGAVVYADRIYVTSQVYSWEVTTAEGGQGLNELLS
KSVLNGIVNGIDINDMNSTDKPLPYHYSVDLSGKAKKELQKELGLPRDPVPL
IGFRLDYKQKIDILIKLAPLMDRNIQFVMLSGDGPFGFEGMRSESGYRDKFRGM
VGFSPVSHRITAGCDIILMPSPFCPCGNQLYAMQYGVVPHVGGTGLRDTVENFP
FAKGGGCTGWAFSPLETKGIADGNFDIQGTQVLLGSNEARHVKRLYMGPCRLTV"
join(45936. .46041,46221. .46516,46748. .46965,47016. .47577,
48123. .48381,50137. .50150)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA81850.1"
/db_xref="GI:5295949"

CDS

Alignment Scores:
Pred. No.: 1,66e-75 Length: 170371
Score: 1355.00 Matches: 280
Percent Similarity: 43.97% Conservative: 91
Best Local Similarity: 34.70% Mismatches: 129
Query Match: 24.12% Indels: 308
Db: 8 Gaps: 8

US-09-830-837-6 (1-1052) x AB026295 (1-170371)

Qy 322 ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341
|||||
Db 166538 GTTTGGAGCTACACAAATAACATTATTATGCTATCAGCTATTGGAATGATGCCCT 166597
Qy 342 LeuTrpGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyIle 361
|||||
Db 166598 CTTTATGACGCTAAATAACCGCTGATCAGAGTGACGCTATTGCTGTGGTGATT 166657
Qy 362 AspPheGluAsnIleAlaArgPheSerArgGlyMetThrThrTrpGluLeuPro 381
|||||
Db 166658 GATTACATAACCATATAGCTTCATTTCTCTCAGGACATGACTACTTGGGAGCTTCT 166717
Qy 381
Db 166718 CATGGGTACATTTCACTTACCAGTACATATTTCTTATGTCTGTCAGTGTGTGT 166777
Qy 381

Db 166778 TTTTTCGATATCTGGATCTCTAGGCTTGAAAAGCTAAAGCTTTGTATCATTTTCATTA 166837
Qy 381
Db 166838 GAACAGAGCTGTACAGTATATAAAACCTCTAAGCCCTAGTTCAGTGGTTTTGTATATATAT 166897
Qy 381
Db 166898 CTTAAATAAATTTGATGGAGTTTGGAGTACTTACATTCGACTTATCCCTATATATATATGT 166957
Qy 381
Db 166958 TGTAAAGGTCATTTGCTAGTTTGTGCTTACCAAACTTGCACCTGCAAGTGCACACAAA 167017
Qy 381
Db 167018 GATGTCATATGCTAATTCCTTTTAAATCATACTAGCAGGTCAAGATAGAAATACGGCTT 167077
Qy 381
Db 167078 ACTGATGTGGTATCATATAAAGCTTTTTTACAAAGATTTCTCCTGAGAAGCTTTGTATAA 167137
Qy 381
Db 167138 TCATATTCGATAATTTTATTTAGTTAGTTGAAATCAGATCTGTTGTGATATTAATTTATGG 167197
Qy 381
Db 167198 CAAGAAACTATTTCTCTGCCCTTGAACCTTTAGGCATTTAATTAATATCTAGTAAAAAC 167257
Qy 381
Db 167258 TTGAATGATGAGCAATTTTTTTGGGGGGGGGGGGGTAGCGATGATCATTTTT 167317
Qy 381
Db 167318 TTACAATGCCATAGTACCATTCTCTTGTGTAGCTTCCACAGATCCTTGAATGGCTTA 167377
Qy 382
Db 167378 CTCATAAGTGTGGTTTCAATAATCCAGG-TATGGCGGTGTAAACCTGATGTTGTGGCG 167436
Qy 394 TyrGlyAlaGlyValArgGlySerGlyVallyGlyCysArgAlaLeuSerGlyThr 413
|||||
Db 167437 TATAGTCGTATATATGGTTCAAGATCAGACAGGTGTGTAACACCTTTTCAGGCACC 167496
Qy 414 SerValAlaSerProValValAlaGlyAlaValThrLeuValSerThrVal- 431
|||||
Db 167497 AGTGTGCAAGCCCGGTGGTGTGCTGTGTAGTGTGTGTTAGTGTATACCAAGAG 167556
Qy 432 --GlnLysArgGluLeuValAsnProAlaSerMetLysGlnAlaLeuAlaSerAla 450
|||||
Db 167557 GAGCATCGAAATCCATCCTTAACCTGCAACAATGAAACAAAGCCCTTGTAGAGGGTCT 167616
Qy 451 ArgArgLeuProGlyValAsnMetPheGluGlnGlyHisGlyLysLeuAspLeu- 468
|||||
Db 167617 TCTAGGCTTTCAGGACCAACATGATGAGCAGGCGCTGGCAAGATTGACCTGTAAGTT 167676
Qy 468
Db 167677 TCTTGTCCCTTTTAAATGTACTTTTGGTATGCCCTATGATTAGTTGGATATTCATACTTC 167736
Qy 468
Db 167737 TTGTATCCCATCTTTTCCAAACGACAGGTTTGTCTTCAAAATTTAGAGCAGTATATC 167796
Qy 468
Db 167797 CTGTCAAGTGTGTTGCCCTTCTTTCATTATGTGTCTCAACAGATACATTTATTTGTTT 167856
Qy 469
Db 167857 TTGTCTTTTCCAGCTGGCAATCATACGAAATCTTGAAAGATTTACCAACCGCTGCAAGC 167916


```
QY 484 LeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSerGln 503
   ::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
167917 ATCTTCCATACCCCTGACTCTCACTGATGTGCTTATCTTGGCTTTTGGCCGCA 167976
QY 504 ProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsnGlyMetGly 523
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
167977 CTTCTGTATGCTGGAGCTATGCGAGTGGTCTCAATGCTACAACTTGAACGGATGGG 168036
QY 524 ValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGlnAsnGlyAsp 543
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
168037 GTGATTGGTTATGTTAAGGATCTCTCTGTATGCGAGCCC--TCTGAGGATGTCGCAAT 168093
QY 544 AsnIleGluValAlaPheSerTyrSerSerValLeuThrProTrpSerGlyTyrLeuAla 563
   ::: ||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
168094 ATTCTTAGTGTTCACCTTCACTTCACTGATGTCATATGCTTGGACAGGTTATCTTGC 168153
QY 564 IleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisVal 583
   ::: ||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
168154 CTGCACCTTACAGTTAAGATGAAGATCTCAGTTTTCAGGCATATATCAGTGGCAAGTTC 168213
QY 584 MetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThrSer 603
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
168214 ACATTGCTCAATTACAGCCGACGA-----GCCATGGGAAAGCAGTCCACGAGTACT 168267
QY 604 ThrValLysLeuProIleLysValLysIleIleProThrProProArgSerLysArgVal 623
   ::: ||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
168268 TCATGTGTTCTTACCTGAGGTCAAGTGTGTCGCAACCTGTAGTCCAGAGAATA 168327
QY 624 LeuTrpAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArgAspAsnLeu 643
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
168328 TTGTGGGACCAATTTCAATATCAATACCACTGCTGGTTTGTTCACAGAGATTCACAT 168387
QY 644 ArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgasp 663
   ::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
168388 AATGTGCATATGATATCTTGTGTCGATGGCATGTCACCTGCACACGATTTTCATATA 168447
QY 664 MetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThr 683
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
168448 CTGTTCAACATCTGTAGATCCAGATACATCTAGACACCTTGGATCTCCACTTACC 168507
QY 684 CysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluTyrPhe 703
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
168508 TCGTTTGATGCTAGCAATATATGGACACTGCTTATGTTGACCTTGAGGATGAGTACTTC 168567
QY 704 ProGluGluIleAlaLysLeuArgArgasp---ValAspAsnGlyLeuSerLeuValIle 722
   ||||| ::::: ||||| ::::: ||||| ::::: |||||
168568 AGTGAGGAGATTTCAGAACTTAAGGATGATGTTGTACACAAAGGGCTTGGCTGTGTT 168627
QY 723 PheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThr 742
   ||||| ::::: ||||| ::::: ||||| ::::: |||||
168628 TTTGCAGAGTGTATCATCTCCACACAATGGTTAAGATGACATCTTTTGTATGAAATACA 168687
QY 743 ArgGlnTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLeu 762
   ||||| ::::: ||||| ::::: ||||| ::::: |||||
168688 CCAGTTGGTGACTCCCAATTAAGGGGTGCAATGTTCCGACATTAATGAACATTTTG 168747
QY 763 SerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyLysGluLeuAlaAsn 782
   ::::: ||||| ::::: ||||| ::::: |||||
168748 GCACCATTTGGCATGCTTTTGGGACAAAGTACTAAGTGGTGACTTCTCAATCAATGCT 168807
QY 783 HisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValVal 802
   ::::: ||||| ::::: ||||| ::::: |||||
168808 GAGCAGACCCACTATGCTTCTGGGACTGATATAGTCAATTCGCCAGCAGGAGTTTGTG 168867
QY 803 IleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGlu-----Thr 819
   ::::: ||||| ::::: |||||
168868 -----CATAGTCTCAATCCAGGACAACTCTAAATTTCTCAGGATAACTCAAGAGC 168921
QY 820 AlaValValGluAsnValPro 826
   ||||| ::::: |||||
168922 GCAGATACACAAAATACCCCA 168942
Db
```

RESULT 18
AP002536
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,
BAC clone:OSJNBa0015114, complete sequence.
ACCESSION
AP002536
VERSION
AP002536.1
KEYWORDS
GI:8698574
SOURCE
HTG.
ORGANISM
Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:OSJNBa0015114.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC
clone:OSJNBa0015114
Published Only in Database (2000)
2 (bases 1 to 105815)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (21-JUN-2000) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
The orientation of the sequence is from -21M13 to M13rev of the BAC
clone.
FEATURES
source
1. .105815
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="6"
/clone="OSJNBa0015114"
BASE COUNT 28862 a 24586 c 23768 g 28599 t
ORIGIN
Alignment Scores: 2,34e-66 Length: 105815
Pred. No.: 1208,50 Matches: 227
Score: 1208,50
Percent Similarity: 60,74% Conservative: 87
Best Local Similarity: 43,91% Mismatches: 124
Query Match: 21,52% Indels: 79
DB: 8 Gaps: 7
US-09-830-837-6 (1-1052) x AP002536 (1-105815)
QY 384 TyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGlySerGlyVal 403
||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 585 TATGGCCGTGTTAAACCTGATGTTGCTGATAGTACGTGATATAATGGTTCAAAGATC 644
QY 404 LysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValValAlaGlyAla 423
||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 645 AGCACAGTTGTAAACCCCTTTCAGCACCAAGTGTTCAGCCGCGGTGCTGTGTGTA 704
QY 424 ValThrLeuValSerThrVal-----GlnLysArgGluLeuValAsnProAla 440
||| ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 705 GTATGCTTCTGTTAGTGTATTATACAGGAGGACATCGGAATTCATCTTACCTTGA 764
QY 441 SerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGlu 460
||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 765 ACAATGAACAAGCCCTTGTAGAGGTGCTTCTAGGCTTTCAGGACCAACAATGATGAG 824
QY 461 GlnGlyHisGlyLysLeuAspLeu----- 468
||||| ||||| ::::: |||||
Db 825 CAGGCGCTGGCAAGATTGACCTGTAAAGTTTCTTGTCCCTTTTAATGTACTCTTGTGAT 894
QY 468 ----- 468
Db 885 GCCTATGATTAGTTGGATATTCAATACTTCTTGTATCCCATCTCTTCCAAACAGCAAGGT 944

QY	468	-----	468
Db	945	TTGTCTTCAAAATTTAGAGCAGTATATCTGTCAGTGGTTGGTTCCTTCCTATT	1004
QY	469	-----	473
Db	1005	GTGTCTCAACAGACTCAATTTATTGTTGTTTCTTTTCCAGCTGGCAATACAGAA	1064
QY	474	IleLeuAsnSerTyrlsProGlnAlaSerLeuSerProSerTyrlsLeuAspLeuThrGlu	493
Db	1065	ATCTTGAAGTTACCAACCGCTGCAAGCATTTCTCTTAATACCTTGACTTCATCAT	1124
QY	494	CysProTyrlsMetTrpProTyrlsCysSerGlnProIleTyrlsGlyMetProThrVal	513
Db	1125	TGTCCCTATTTCTGGCTTTTGGCGCAACCTCTCTATGCTGGAGTATGCCAGTGGTC	1184
QY	514	ValAsnValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAsp	533
Db	1185	TTCATGCTACAACTCTGAACGGGATGGGTGATTGGTTATGTTAAGGATCTCTCTGTA	1244
QY	534	TrpGlnProTyrlsLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrlsSer	553
Db	1245	TGGCAGCCC---TCTGAGGATGTCGGCAATATCTTAGTGTCTACTTCACTTACGAT	1301
QY	554	ValLeuTrpProTrpSerGlyTyrlsAlaIleSerIleSerValThrLysLysAlaAla	573
Db	1302	GTATATGCTTGGCAGGTATCTTCCCTGACACTTACAGTTAAGATGAAGCACT	1361
QY	574	SerTrpGluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThr	593
Db	1362	CAGTTTTCAGGCATAATCAGTGGCAAGTGCACATGTCATTTTACAGCCAGCA	1415
QY	594	GluSerLysAsnGlyAlaGluGlnThrSerThrValLysLeuProIleLysValLysIle	613
Db	1416	GCCATGCGGGAAGCAGTCCACGAAGTAGTTCATGTTCTTACCTGAAGGTCAGGTG	1475
QY	614	IleProTrpProProArgSerLysArgValLeuTrpAspGlnTyrlsAsnLeuArgTyrl	633
Db	1476	GTCCCAACACCTGTAGTCCAGAAAGATATTGTGGGACCAATTTTCATATATCAATAC	1535
QY	634	ProProGlyTyrlsPheProArgAspAsnLeuArgMetLysAsnAspProLeuAspTrpAsn	653
Db	1536	CCATCTGCTTTTGTCCAGAGATTCATTAATGTGCATAATGATATCTTACTGGCAT	1595
QY	654	GlyAspHisIleHisThrAsnPheArgAspMetTyrlsGlnHisLeuArgSerMetGlyTyrl	673
Db	1596	GCGCATCATTGCACAGATTTTCATATATCTGTCAACATCTGAGAGATGAGGATAC	1655
QY	674	PheValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrlsLeu	693
Db	1656	TACTTGGACCTTGGATCTCCACTTACCTTGTGCTTGTGCTGCTGCTGCTGCTGCTG	1715
QY	694	LeuMetValSerGluGluGluTyrlsPheProGluIleAlaLysLeuArgArgAsp	713
Db	1716	CTTATGTTGACCTTGGAGTACGATCTTACGAGGAGATTCAGAACTTAAGGATGAT	1775
QY	714	---ValAspAsnGlyLeuSerLeuValIlePheSerAspTrpTyrlsAsnThrSerValMet	732
Db	1776	GTGTGTACAAAGGGCTGGCGTTGTTGTTTTCAGAGTGTATCATCTGCGACAAATG	1835
QY	733	ArgLysValLysPheTyrlsAspGluAsnThrArgGlnTrpMetProAspThrGlyGly	752
Db	1836	GTTAAGATGACATCTTTTGTGAAATACACGAGTGTGGTGGTCAATTAAGTGGGGT	1895
QY	753	AlaAsnIleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGly	772
Db	1896	GCAATGTTCTGCACTTAATGAACCTTTTGGCACCATTGGCATTTGCTTTGGGCAAA	1955
QY	773	LeuTyrlsGluGlyGluPheThrLeuAlaAsnHisAspMetTyrlsAlaSerGlyCysSer	792
Db	1956	GTACTAAGTGTGACTTCTCAATCAATGTTGAGCAGACCCACTATGCTTCTGGGACTGAT	2015
QY	793	IleAlaLysPheProGluAspGlyValIleThrGlnThrPheLysAspGlnGlyLeu	812

Db	2016	ATAGTCAATTCACAGAGGTTTGTG-----CATAGCTTCCACTCCAGCAAC	2069
QY	813	GluValLysGlnGlu-----ThrAlaValValGluAsnValPro	826
Db	2070	TCTAAATTTCTCAGGATAACTCAAGAGCGCAGATACACAAATATCCCA	2120
RESULT 19			
AC010662		113320 bp	DNA linear HTG 18-FEB-2000
LOCUS			
DEFINITION			Drosophila melanogaster clone RPC198-2K23, *** SEQUENCING IN
ACCESSION			PROGRESS ***, 66 unordered pieces.
VERSION			AC010662.5 GI:6996787
KEYWORDS			HTG; HTGS_PHASE1
SOURCE			Drosophila melanogaster
ORGANISM			Drosophila melanogaster
REFERENCE			Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS			1. (bases 1 to 113320) Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K., Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C., Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C., David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Frantz, P., Ganes, R., Ganes, R., Gorrell, J.H., Gorrell, L.L., Guvara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M., Hollway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondejowski, N., Kong, Y., Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, R., Nguyen, S., Osval, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R., Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wabnah, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.
TITLE			Direct Submission
JOURNAL			Unpublished
REFERENCE			2. (bases 1 to 113320) Worley, K.C.
AUTHORS			Direct Submission
JOURNAL			Submitted (17-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT			On Feb 18, 2000 this sequence version replaced gi:6056120. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc.help@bcm.tmc.edu ----- Project Information Center project name: DRG8-2K23 Center clone name: RPC198-2K23 ----- Summary Statistics Sequencing vector: ML3; L08821 Chemistry: Dye-terminator Big Dye; 41% of reads Assembly program: Phrap; version 0.980611 Consensus quality: 52892 bases at least Q40 Consensus quality: 72110 bases at least Q30 Consensus quality: 84532 bases at least Q20 Estimated insert size: 100525; sum-of-contigs estimation Quality coverage: 1.2x in Q20 bases; sum-of-contigs estimation ----- * NOTE: This is a 'working draft' sequence. It currently consists of 66 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	984:	contig of 984 bp in length	40056	41772:	contig of 1717 bp in length
985	1004:	gap of unknown length	41773	41792:	gap of unknown length
1005	2123:	contig of 1119 bp in length	41793	43162:	contig of 1370 bp in length
2124	2143:	gap of unknown length	43163	43182:	gap of unknown length
3254	3254:	contig of 1111 bp in length	43183	44086:	contig of 904 bp in length
3255	3274:	gap of unknown length	44087	44106:	gap of unknown length
3275	4526:	contig of 1252 bp in length	44107	45469:	contig of 1363 bp in length
4527	4546:	gap of unknown length	45470	45489:	gap of unknown length
4547	5431:	contig of 885 bp in length	45490	47159:	contig of 1670 bp in length
5432	5451:	gap of unknown length	47160	47179:	gap of unknown length
5452	5936:	contig of 485 bp in length	47180	48716:	contig of 1537 bp in length
5937	5956:	gap of unknown length	48717	48737:	gap of unknown length
5957	7052:	contig of 1096 bp in length	48738	50174:	contig of 1438 bp in length
7053	7072:	gap of unknown length	50175	50194:	gap of unknown length
7073	7889:	contig of 817 bp in length	50195	51756:	contig of 1562 bp in length
7890	7909:	gap of unknown length	51757	51776:	gap of unknown length
7910	9199:	contig of 1290 bp in length	51777	53606:	contig of 1830 bp in length
9200	9219:	gap of unknown length	53607	53627:	gap of unknown length
9201	10004:	contig of 785 bp in length	53628	55470:	contig of 1844 bp in length
10005	10024:	gap of unknown length	55471	56632:	contig of 1142 bp in length
10025	10988:	contig of 974 bp in length	56633	56652:	gap of unknown length
10989	11018:	gap of unknown length	56653	59120:	contig of 2468 bp in length
11019	11801:	contig of 783 bp in length	59121	59140:	gap of unknown length
11802	11821:	gap of unknown length	59141	60514:	contig of 1374 bp in length
11822	12546:	contig of 725 bp in length	60515	60534:	gap of unknown length
12547	12566:	gap of unknown length	60535	61864:	contig of 1330 bp in length
12567	13350:	contig of 784 bp in length	61865	61884:	gap of unknown length
13351	13370:	gap of unknown length	61885	64010:	contig of 2126 bp in length
13371	14218:	contig of 848 bp in length	64011	64030:	gap of unknown length
14219	14238:	gap of unknown length	64031	65825:	contig of 1795 bp in length
14239	15403:	contig of 1165 bp in length	65826	65845:	gap of unknown length
15404	15423:	gap of unknown length	65846	67721:	contig of 1876 bp in length
15424	16876:	contig of 1453 bp in length	67722	70003:	contig of 2262 bp in length
16877	16896:	gap of unknown length	70004	70023:	gap of unknown length
16897	18099:	contig of 1203 bp in length	70024	72638:	contig of 2615 bp in length
18100	18119:	gap of unknown length	72639	72658:	gap of unknown length
18120	19038:	contig of 919 bp in length	72659	75014:	contig of 2356 bp in length
19039	19058:	gap of unknown length	75015	75034:	gap of unknown length
19059	20658:	contig of 1600 bp in length	75035	77594:	contig of 2560 bp in length
20659	20678:	gap of unknown length	77595	77614:	gap of unknown length
20679	21926:	contig of 1248 bp in length	77615	80837:	contig of 3223 bp in length
21927	21946:	gap of unknown length	80838	80857:	gap of unknown length
21947	23095:	contig of 1149 bp in length	80858	83808:	contig of 2951 bp in length
23096	23115:	gap of unknown length	83809	83828:	gap of unknown length
23116	24605:	contig of 1490 bp in length	83829	85733:	contig of 1905 bp in length
24606	24625:	gap of unknown length	85734	85753:	gap of unknown length
24626	25352:	contig of 727 bp in length	85754	88139:	contig of 2386 bp in length
25353	25372:	gap of unknown length	88140	88159:	gap of unknown length
25373	27203:	contig of 1831 bp in length	88160	91070:	contig of 2911 bp in length
27204	27223:	gap of unknown length	91071	91090:	gap of unknown length
27224	28268:	contig of 1045 bp in length	91091	95700:	contig of 4610 bp in length
28269	28288:	gap of unknown length	95701	95720:	gap of unknown length
28289	29754:	contig of 1466 bp in length	95721	99302:	contig of 3582 bp in length
29755	29774:	gap of unknown length	99303	99322:	gap of unknown length
29775	31341:	contig of 1567 bp in length	99323	103945:	contig of 4623 bp in length
31342	31361:	gap of unknown length	103946	103965:	gap of unknown length
31362	32418:	contig of 1057 bp in length	103966	108234:	contig of 4269 bp in length
32419	32438:	gap of unknown length	108235	108254:	gap of unknown length
32439	33457:	contig of 1019 bp in length	113320	contig of 5066 bp in length.	
33458	33477:	gap of unknown length			
33478	34737:	contig of 1260 bp in length			
34739	34757:	gap of unknown length			
34758	35960:	contig of 1203 bp in length			
35961	35980:	gap of unknown length			
35981	37758:	contig of 1778 bp in length			
37759	37778:	gap of unknown length			
37779	38601:	contig of 823 bp in length			
38602	38621:	gap of unknown length			
38622	40035:	contig of 1414 bp in length			
40036	40055:	gap of unknown length			

FEATURES source

Alignment Scores:

Pred. No.: 2.95e-43 Length: 113320
 Score: 850.50 Matches: 173
 Percent Similarity: 64.00% Conservatives: 67
 Best Local Similarity: 46.13% Mismatches: 96
 Query Match: 15.14% Indels: 44
 DB: 2 Gaps: 4

US-09-830-837-6 (1-1052) x AC010662 (1-113320)

QY	901	-----	901	-----	901
Db	997	GAACATGATCTACTTCAGCCAGTGGCTGTGCTGCCAGCAAGTGGGGCCAGTTTGTTG	2077	TAGACACAGGGCTCTGTCTGTCTGTGGAAAGACAGACAGACTCATATAAGCCCT	2136
QY	901	-----	901	-----	901
Db	1057	CGCGCTGAGTGGTTTTTCGTAATGCTGCGGTAGAGCAGCACTCAGCGGCTCTAACCTG	2137	CTTTTTCTTTTCATTTTAGTAATAGAGATGATTTCTTTTGTGTGCTGTGCTA	2196
QY	901	-----	901	-----	901
Db	1117	CCGAGAGGCCATCCTTAGTCCCGAGAGGCCATGATCGAGGACCCAGTGTGCAGCTGCTT	902	-----GlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspPro	919
QY	901	-----	901	-----	901
Db	1177	TCCCACGTGCACCCGACCCTCAGCCACTCTGAGTGTGTTTACCCTCCGCTCTCACAG	2197	ACCACAGAAACCACTTCATCGGTACTCCAAGGTTCTGGAGGCCCACTTTGGGAGACCA	2256
QY	901	-----	901	-----	901
Db	1237	GTGAGGACAGTGGCTCCTACAGATCTTTCAGGAAGAAATCACCACATTTCTAGAGCACTGC	920	LysProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsn	939
QY	901	-----	901	-----	901
Db	1297	CAGTCCCATGACAGAGCTGGGATGCTCCATGGATTACCTCACTGTCCCTCTCAGTA	2257	AAACCTCGGCTCTACACAGCTGTCCAGCTGTCTTGGGCCAAGCCACACACCTTTAAAC	2316
QY	901	-----	901	-----	901
Db	1357	ATCCTGTGAAATAGATGTTATTATTTCCAGTGTTCAGATGAGGAATGAGAGCTGAGAG	940	GluThrAlaProSerAsnLeuTrpLysHisGlnLysLeuSerIleAspLeuAspLys	959
QY	901	-----	901	-----	901
Db	1417	AAATTAGATAACTGCCTGAGATGCCTGGCGGTATGTGTGTGGAGCCTCAACCGGCC	2317	GAGACGGCGCCAGTAACTTTGGAACATCAGAAGCTACTCTCATTTGACCTGGACAAG	2376
QY	901	-----	901	-----	901
Db	1477	TCGTGGGCGCTGTCTGTGGTGGCGGCCAGCTGTCTGGCACATTAGTGGGGTGGTA	960	ValValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGlu	979
QY	901	-----	901	-----	901
Db	1537	TCAGTGGCTAAGTCCCTGGTGTGCTGGTTCATAGCGCTGCCCTAGAACGCTCATG	2377	GTGGTGTATACCACTTTTCGATCGAATGCGCCCTCAAGTGAGGCCCTTTGCCCTGGAGAG	2436
QY	901	-----	901	-----	901
Db	1597	GGAACACAGGCATCATTTAAATTTTCCAGAGCTGCGTTTAAAAAGTAAAAAGAAACA	980	SerGlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGly	999
QY	901	-----	901	-----	901
Db	1657	CAGATGAACATATTTTAATAATATCTTGAGCCCAAGACATGAAGAATATTAGAAATTAA	2437	AGCGGCGCTTGGGACATCTCTGGAGGATCATGCTGGCGCTACACAGGAGGTGGGC	2496
QY	901	-----	901	-----	901
Db	1717	ACAGGTGATACCTATTTCACCTGCTCAGTAGCTACGTGTGGTGGCACAATGGTGACA	1000	GlnThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValVal	1019
QY	901	-----	901	-----	901
Db	1777	TCTAAGAAATAGAAAGAAATTTTCAAAATGCGAGAGTGGTTTTTATTATTGGCCAGAAATTA	2497	CAGACCATCTCTGCTTCTTGGCTTCTCTGGAGCCATGGTGTCTCTGGCGCTTCTTGTGTGTA	2556
QY	901	-----	901	-----	901
Db	1837	AATATTTTAGAGGGAAGTATTTTTTCCATGAAACCGGTCTGTAGCTTAGATTTCGCAGG	1020	GlnIleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGln	1039
QY	901	-----	901	-----	901
Db	1897	TGTAGCTTCCGTGTCTCAGTGTGTGTGTCAGCAAAAGTTTGACAGCCAGCTGCTGTGCACAC	2557	CAATCAACAGGCCAAGACAGCGCCGAGGAGAGCCAGGGTGAAGCGCCCGCAG	2616
QY	901	-----	901	-----	901
Db	1957	ATCTGCCCATCTTCAAGGGAACGGTCTGTGATTTTTCACCCACCTTCAGCAATGGAACA	1040	LeuMetGlnGlnValHisProLysThrProSerVal	1052
QY	901	-----	901	-----	901
Db	2017	CACATTGCGGGGAGGAACCAATCAGCGGTGGCTCCTGGGCTCATGGAGGATAGCTTAA	2617	CTCATGACGAGGTTCACCGCCCAAGACCCCTTCGGGTG	2655

RESULT 21
AC010056
LOCUS
DEFINITION
AC010056 129955 bp DNA linear HTG 16-OCT-1999
SEQUENCING IN PROGRESS ***, 1 unordered piece.
AC010056 GI:6056152
HTG; HTGS_PHASE1.
SOURCE
Drosophila melanogaster.
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 129955)
AUTHORS
Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,
Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
David, R., Delgado, O., Deshazo, D., Ding, F., Domah-Rashid, N.,
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
Forum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M.,
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,
Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
Lichtharge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R.,
Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S.,
Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswal, G.,
Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M.,
Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E.,
Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R., Tabor, P.,
Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M., Watlington, S.,
Weinstock, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J.,
Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.
Direct Submission

TITLE


```

QY 682 eThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGluTyr 702
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 18159 CACCTGCTCAAGTCCGATGATGCGCGGTATTGATGTTGACCCNTGAGAGGGGT 18218
*
QY 702 rPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeuValIle 722
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 18219 TTGGCAGCAGGAGGACAAAGCTGCTTACGATG--GACAATGCATACGACTTGCT 18276
*
QY 722 ePheSerAspTyrTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnTh 742
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 18277 ATTAGCAGTACGTTTATCATCCCTGTTTATGTAATGTCATGGGAGGAGCAACAAC 18336
*
QY 742 rArgGlnTyrTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuL 762
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 18337 AATAAGAGGCTTCTCGCCAGGACGACAGATTTATGGCCGCCCAATATGTCAAAATATT 18396
*
QY 762 euSerValTrpAsnMetGly 768
||| |||||
Db 18397 AAGAGACATGGACAATGGC 18416
*
RESULT 22
AC011909
LOCUS AC011909 129955 bp DNA linear HTG 16-OCT-1999
DEFINITION Drosophila melanogaster chromosome 3L/79A3 clone RPC198-25D3, ***
SEQUENCING IN PROGRESS ***, 1 unordered piece.
ACCESSION AC011909
VERSION AC011909.1 GI:6056127
KEYWORDS HTG; HTGS_PHASEL.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 129955)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejowski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabar,P.,
Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wabnah,M., Watlington,S.,
Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
Wrenstock,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 129955)
Worley,K.C.
Direct Submission
Submitted (16-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 820: contig of 820 bp in length
* 821 1892: contig of 1072 bp in length
* 1893 3138: contig of 1246 bp in length
* 3139 4261: contig of 1123 bp in length
*
FEATURES
source
1..129955
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L/79A3"
/clone="RPC198-25D3"
BASE COUNT 35760 a 28458 c 29103 g 36584 t 50 others
ORIGIN
Alignment Scores: 2.8e-38 Length: 129955
Pred. No.: 774.50 Matches: 163
Score: 59.28% Conservative: 67
Percent Similarity: 42.01% Mismatches: 118
Best Local Similarity: 13.79% Indels: 42
Query Match: 2 Gaps: 4
DB:
US-09-830-837-6 (1-1052) x AC011909 (1-129955)
QY 417 SerProValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGlu 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 17272 TCCTCCAGTTGTTGCGAGGGCTGCTGCTTATAGCGTGCATTTCAGAAATCGAC 17331
*
QY 436 LeuValAsnProAlaSerMetLysGlnAlaLeuAlaSerAlaArgLeuProGly 455
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Db	17332	TACATAAACCCAGCATCTCTTAAGCAGGTACTCATTTGAAGTGGCGGAGAAATCGCCGAT	17391
QY	456	ValAsnMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeu	475
Db	17392	TATAACATGTTTGGCAGGAGCTGGAAACCTGAATTCGTGAAGAGTATGCGATTTG	17451
QY	476	AsnSerTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysPro	495
Db	17452	CTGTCATACAAACCAAGATAACCTTATTCGGGCATACCTTGACTTCACCAA---AAC	17508
QY	496	TyrMetTrpProTyrCysSerGlnProIleTyrTyrGlyMetProThrValValAsn	515
Db	17509	TATATGGGCTTATAGTCCCAACCTCTGCTACTATGGAACCTCCGTCGCTATTCACAA	17568
QY	516	ValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGln	535
Db	17569	GTTACCATCTCAATGTTATCTCTGTCACAGTATAGTTCGATCCCTTAATGGATT	17628
QY	536	ProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeu	555
Db	17629	CCCGATTTCGAAACCAAGGTCAGTTCTTCAAGTATCTGCACAAAGTTTCGCTATTCGT	17688
QY	556	TrpProTrpSerGlyTyrLeuAlaIleSerIle-----	566
Db	17689	TGGCGGTGGCGGTTGGATGTCAGTTTATTTGTTGAGTATGAAATAGCAACGCTAT	17748
QY	566	-----	566
Db	17749	ATAAAGATATATGGGTACATATTTGATATGATAAGTGTCTAAGCCATCTTATACCAACAA	17808
QY	567	-----Ser-ValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHI	582
Db	17809	TTTTTTTACACGCTGTAAAGAGGAGGAGAAACTTGAAGGTGTTTGAAGAGGAG	17868
QY	582	svAlMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluTh	602
Db	17869	TATCACCCTAGTTTGTGAAGCTTTTAAACAGACACCACCAACGAACT-----CATGT	17919
QY	602	rSerThrValLysLeuProIleLysValLysIleIleProThrProArgSerLysAr	622
Db	17920	TACAGAGTCGACTTCTTAAACATAAAGGTACT-CCAAACCGCCAGGAACAGAG	17978
QY	622	gValLeuTrpAspGluTyrHisAsnLeuArgTyrProGlyTyrPheProArgAspAs	642
Db	17979	GATTTTATGGATGATGATACACAGCTTAAGTATCCACCGCGCTATATCCAGAGATGA	18038
QY	642	nLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheAr	662
Db	18039	TCTCAAAGTTAACTAGATCCTCTGGACTGGAGGCGGAGACCATATACACAACTTTAG	18098
QY	662	gAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPh	682
Db	18099	GGACATGTATACATTTACGAAATGTTGGCTACTATGATGTTTGGGAGAACCTTT	18158
QY	682	eThrCysPheAspAlaSerGluTyrGlyThrLeuLeuMetValAspSerGluGluGluTyr	702
Db	18159	CACCTGCTCAATCCCTCGGATTATGGCGGCTTATGATTGTTGACCCNTGAGAGAGGT	18218
QY	702	rPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeuValIle	722
Db	18219	TTGGCGCGCAGGACATAAAGCTGCTTACGATG-GACAATGATGACGACTTGTCT	18276
QY	722	ePheSerAspTrpTrpAsnThrSerValMetArgLysValLysPheTyrAspGluAsnTh	742
Db	18277	ATTAGCCAGTACGTTTCATCCTGTTTATGTAAGATGTCATGGGAGGAGGAGCAACAC	18336
QY	742	rArgGlnTrpTrp-MetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuL	762
Db	18337	AATAGAGGCTTCTCGGCCAGGACCATGATTTGGCGCGCCCAATATGTCAAAATATT	18396
QY	762	euSerValTrpAsnMetGly 768	
Db	18397	AAGAGACATGGCAATGGGC 18416	

RESULT 23	AC040169/c	214287 bp	DNA	linear	HTG 12-JAN-2002
LOCUS	AC040169	214287 bp	DNA	linear	HTG 12-JAN-2002
DEFINITION	AC040169	214287 bp	DNA	linear	HTG 12-JAN-2002
ACCESSION	AC040169	214287 bp	DNA	linear	HTG 12-JAN-2002
VERSION	AC040169.6	214287 bp	DNA	linear	HTG 12-JAN-2002
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.	214287 bp	DNA	linear	HTG 12-JAN-2002
SOURCE	Homo sapiens	214287 bp	DNA	linear	HTG 12-JAN-2002
ORGANISM	Homo sapiens	214287 bp	DNA	linear	HTG 12-JAN-2002
REFERENCE	DOE Joint Genome Institute.	214287 bp	DNA	linear	HTG 12-JAN-2002
AUTHORS	DOE Joint Genome Institute.	214287 bp	DNA	linear	HTG 12-JAN-2002
TITLE	Sequencing of Human Chromosome 16	214287 bp	DNA	linear	HTG 12-JAN-2002
JOURNAL	Unpublished	214287 bp	DNA	linear	HTG 12-JAN-2002
REFERENCE	DOE Joint Genome Institute.	214287 bp	DNA	linear	HTG 12-JAN-2002
AUTHORS	DOE Joint Genome Institute.	214287 bp	DNA	linear	HTG 12-JAN-2002
TITLE	Direct Submissio	214287 bp	DNA	linear	HTG 12-JAN-2002
JOURNAL	Submitted (11-APR-2000) Production Sequencing Facility, DOE Joint	214287 bp	DNA	linear	HTG 12-JAN-2002
COMMENT	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	214287 bp	DNA	linear	HTG 12-JAN-2002
	On Jan 12, 2002 this sequence version replaced gi:13786410.	214287 bp	DNA	linear	HTG 12-JAN-2002
	Sequence Quality Assessment:	214287 bp	DNA	linear	HTG 12-JAN-2002
	This entry has been annotated with sequence quality	214287 bp	DNA	linear	HTG 12-JAN-2002
	estimates computed by the Phrap assembly program.	214287 bp	DNA	linear	HTG 12-JAN-2002
	All manually edited bases have been reduced to quality zero.	214287 bp	DNA	linear	HTG 12-JAN-2002
	Quality levels above 40 are expected to have less than	214287 bp	DNA	linear	HTG 12-JAN-2002
	1 error in 10,000 bp.	214287 bp	DNA	linear	HTG 12-JAN-2002
	Base-by-base quality values are not generally visible from the	214287 bp	DNA	linear	HTG 12-JAN-2002
	GenBank flat file format but are available as part	214287 bp	DNA	linear	HTG 12-JAN-2002
	of this entry's ASN.1 file.	214287 bp	DNA	linear	HTG 12-JAN-2002
	-----	214287 bp	DNA	linear	HTG 12-JAN-2002
	Sequence Quality Assessment:	214287 bp	DNA	linear	HTG 12-JAN-2002
	This entry has been annotated with sequence quality	214287 bp	DNA	linear	HTG 12-JAN-2002
	estimates computed by the Phrap assembly program.	214287 bp	DNA	linear	HTG 12-JAN-2002
	All manually edited bases have been reduced to quality zero.	214287 bp	DNA	linear	HTG 12-JAN-2002
	Quality levels above 40 are expected to have less than	214287 bp	DNA	linear	HTG 12-JAN-2002
	1 error in 10,000 bp.	214287 bp	DNA	linear	HTG 12-JAN-2002
	Base-by-base quality values are not generally visible from the	214287 bp	DNA	linear	HTG 12-JAN-2002
	GenBank flat file format but are available as part	214287 bp	DNA	linear	HTG 12-JAN-2002
	of this entry's ASN.1 file.	214287 bp	DNA	linear	HTG 12-JAN-2002
	-----	214287 bp	DNA	linear	HTG 12-JAN-2002
	* NOTE: This is a 'working draft' sequence. It currently	214287 bp	DNA	linear	HTG 12-JAN-2002
	* consists of 4 contigs. Gaps between the contigs	214287 bp	DNA	linear	HTG 12-JAN-2002
	* are represented as runs of N. The order of the pieces	214287 bp	DNA	linear	HTG 12-JAN-2002
	* is believed to be correct as given, however the sizes	214287 bp	DNA	linear	HTG 12-JAN-2002
	* of the gaps between them are based on estimates that have	214287 bp	DNA	linear	HTG 12-JAN-2002
	* provided by the submitter.	214287 bp	DNA	linear	HTG 12-JAN-2002
	* This sequence will be replaced	214287 bp	DNA	linear	HTG 12-JAN-2002
	* by the finished sequence as soon as it is available and	214287 bp	DNA	linear	HTG 12-JAN-2002
	* the accession number will be preserved.	214287 bp	DNA	linear	HTG 12-JAN-2002
	* 1 9563: contig of 9563 bp in length	214287 bp	DNA	linear	HTG 12-JAN-2002
	* 9564 9663: gap of unknown length	214287 bp	DNA	linear	HTG 12-JAN-2002
	* 9664 26317: contig of 16654 bp in length	214287 bp	DNA	linear	HTG 12-JAN-2002
	* 26318 42789: contig of unknown length	214287 bp	DNA	linear	HTG 12-JAN-2002
	* 26418 42789: contig of 16372 bp in length	214287 bp	DNA	linear	HTG 12-JAN-2002
	* 42790 42889: gap of unknown length	214287 bp	DNA	linear	HTG 12-JAN-2002
	* 42890 214287: contig of 171398 bp in length.	214287 bp	DNA	linear	HTG 12-JAN-2002
	* Location/Qualifiers	214287 bp	DNA	linear	HTG 12-JAN-2002
	1. 214287	214287 bp	DNA	linear	HTG 12-JAN-2002
	/organism="Homo sapiens"	214287 bp	DNA	linear	HTG 12-JAN-2002
	/db_xref="taxon:9606"	214287 bp	DNA	linear	HTG 12-JAN-2002
	/chromosome="16"	214287 bp	DNA	linear	HTG 12-JAN-2002
	/clones="RP11-505K9"	214287 bp	DNA	linear	HTG 12-JAN-2002
	BASE COUNT 57135 a 53978 c 52644 g 50229 t 301 others	214287 bp	DNA	linear	HTG 12-JAN-2002
	ORIGIN	214287 bp	DNA	linear	HTG 12-JAN-2002
	Alignment Scores:	214287 bp	DNA	linear	HTG 12-JAN-2002
	Pred. No.: 5,55e-35	214287 bp	DNA	linear	HTG 12-JAN-2002
	Score: 728.00	214287 bp	DNA	linear	HTG 12-JAN-2002
	Percent Similarity: 18.54%	214287 bp	DNA	linear	HTG 12-JAN-2002
	Best Local Similarity: 18.01%	214287 bp	DNA	linear	HTG 12-JAN-2002
	Mismatches: 19	214287 bp	DNA	linear	HTG 12-JAN-2002

Db	127203	GGGGCTATTGCTAAGCTCTTCTGTGACACACAGATGGTGGCCAGCAATTTACTTGAAGC	127144
QY	652	-----	652
Db	127143	CGACTGTACAGTGAATGCTCCAGGACACATTAACATGCTTTTGGCCATTTCTCTGGTC	127084
QY	652	-----	652
Db	127083	ATTTTGTCTCTTCATGCACCTGCCAGGCACATGGTTCTTCTTATCAGCCATTTCCGTG	127024
QY	652	-----	652
Db	127023	TGTCAGACAGACACACAGGTCTCTAAGCTGATGCTGTGCGCAGTGGCCAGCCATC	126964
QY	652	-----	652
Db	126963	CAGGCCCTTCTGTGGCCATGGCATGGTCTTGCAGAGCCGCCCTGTGGCCAGATA	126904
QY	653	-----As	653
Db	126903	TGCCCTTATTGCCAGTAAGATCGATCAATTTAATGCTGTGGCCTATTGTTTTCAGGAA	126844
QY	653	nglyAspHisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTy	673
Db	126843	TGGTATCATATPCCACACCAATTTCCAGGATATGTACCAAGCATCTGAGAGCATGGGCTA	126784
QY	673	rPheValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThr--	692
Db	126783	CTTTGTAGAGTCTCTCGGGCCCCCTTCACGTGTTTTGTATGCCAGTCAGTATGGTAA--GT	126725
QY	692	-----	692
Db	126724	GGCACTGTGTGCTAGCCAGATGGTCTTTCCGCCCACTTTGGATGTTGTAAATGCTG	126665
QY	692	-----	692
Db	126664	AGGAGTGAATCCAGTTTTTCAGTGATAGCTACCAAGGAGTATGCGCGTGGCCCCACA	126605
QY	692	-----	692
Db	126604	GCTGGGTGAGCCTGGGCGAGGCTCTGAACAGGTGGTGGCGGTGACATTCAGCTGAA	126545
QY	692	-----	692
Db	126544	ATACAAGACAGGCTGATTGGAGACAGGACATTCGTACTCTTTTCTTTAAGTTAAAG	126485
QY	692	-----	692
Db	126484	CAGCTCTGCTATATTTTGTCTCTCTCAAGTGAATGAATACTTTATCTGAGCACCA	126425
QY	692	-----	692
Db	126424	CAAGTGAATCAGATCAGTTAGTCGCTGTGTATTGCTGGATCCTGCTCTACTTTT	126365
QY	692	-----	692
Db	126364	TAGTCCCTTTTAAAGTGTACAGGTGGCACAACACAGACTGTTAGGTGTCAGTAACCTG	126305
QY	692	-----	692
Db	126304	AGACATGTACTGTGTACACTGTTTCTGCCGTGGTGTGCTGCTGTCATTGAGAAGTGACA	126245
QY	693	-----LeuLeuMetValAspSerGluGluG	701
Db	126244	GTTCGCCAGCTTTGTTTGTTCATCAGGCATTTGCTGATGGTGACAGTGGAGGA	126185
QY	701	uTyrPheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeuVa	721
Db	126184	GTACTTCCCTGAAGATCCCAAGCTCCGGGGAGCGTGGACACAGCCCTCTCGCTCGT	126125
QY	721	lIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAs	741
Db	126125	-----	741
Db	126124	CATCTCAGTGTGCTGACACACACTCTCTGTTATGAGAAAAGTGAAGTCTTTTATGATGAAAA	126065
QY	741	nThr-----	742
Db	126064	CACAAGGTACTGTTGATACGCATTTGTTGGATAAGACTTTTGGGAATAAATCTTA	126005
QY	742	-----	742
Db	126004	CCCAGCAAAATACTTAAAGTGTATTTGACCTGTTGGTGTCTTTCCAGTTGGCACAGACA	125945
QY	742	-----	742
Db	125944	CATTGAGAGCTTTGATTAGAAAAAAGCCCTTTTCGACTGAGTGGCTGTCTCACGCCCTG	125885
QY	742	-----	742
Db	125884	TAATCCAGCACTTTGAAGGCTGAGTTGGGCAGATCGGTTGAACCTGGGAGGCCAAGGC	125825
QY	742	-----	742
Db	125824	TGCAGTGTGAGTGTAGATCAAGCCACTGTACTCCAGCTGGGTGTACAGAGACCCCTATCTC	125765
QY	742	-----	742
Db	125764	AAAGAGAAAAAATAAACAACAACAAAAAATAGTCTCTTTATACAGTAGAGAGTTCTTG	125705
QY	742	-----	742
Db	125704	AAAGTTGCTCATAGAGTAAGATGTTTATTTGAGTATGAGTACTCTTAATAATAATAGAGGT	125645
QY	742	-----	742
Db	125644	TCATGTCTTCTTCTTAACTTTAGATATGAGTACTTTCAGGTTTAAACTTCTGTTTAA	125585
QY	742	-----	742
Db	125584	TTCTTTTCCACTCTCTTACTTACTGATATATAAATAAATGACCAAAAGAACTGTTT	125525
QY	742	-----	742
Db	125524	CITTAAGAATATTTTCAGTAAATTTATGTATATTTAAATATTTATATTACAAATAATTA	125465
QY	742	-----	742
Db	125464	ATTACTTATATTATATAGTAAATTAATTTACTCAAGGGAATATGTCCTTCCCTTATTG	125405
QY	742	-----	742
Db	125404	GGATCAAGGACTCCCAATGTATATAAACATAATCACCTTTATGTTTGAATCGGTTTCT	125345
QY	742	-----	742
Db	125344	TTAGAATCTGTGACTGAACATTTCTATGCTTTTCTCATCTCTTCTCCTTCTACTGAAC	125285
QY	742	-----	742
Db	125284	CTTGCTGGTCTAGGCTTAAGCAGTTATCTCTGGTGCATTTTCGGTAAAGTACTGATTTTCA	125225
QY	742	-----	742
Db	125224	TTTCATAAATAAATTTTGTCTTTAATTAATTTACTACATCAGCCGAAGCACTTAAATTTG	125165
QY	742	-----	742
Db	125164	CAGTTTACCCAAACCCAGACAGTTGGTTTCAGTTGAAGCCTTAGTGTGAAGGACTGGA	125105
QY	742	-----	742
Db	125104	TTGTGGCAGGAGAAATAGTTTGTGTACAGAGACTAAATACTGTGTATGGAGTGTCTT	125045
QY	743	-----ArgGlnTrpTrpMetProAspTh	750
Db	125044	CATTCCCTGTTCCCTGTGTGTCTGTGTACCACCTTTCCCGAGCGAGTGTGTGATCCGGATAC	124985

Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 195287)
 Worley, K.C.
 Direct Submission
 Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 195287)
 Worley, K.C.
 Direct Submission
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20303366.

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: GSSS
 Center clone name: CH230-12D23

 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 137024 bases at least Q40
 Consensus quality: 143926 bases at least Q30
 Consensus quality: 150275 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 65 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1272: contig of 1272 bp in length
 * 1273: gap of unknown length
 * 1373 1372: contig of 1368 bp in length
 * 2741 2840: gap of unknown length
 * 2841 4272: contig of 1432 bp in length
 * 4273 4372: gap of unknown length
 * 4373 5666: contig of 1294 bp in length
 * 5667 7266: contig of 1500 bp in length
 * 7267 7366: gap of unknown length
 * 7367 8912: contig of 1546 bp in length
 * 8913 9012: gap of unknown length
 * 9013 10022: contig of 1010 bp in length
 * 10023 10122: gap of unknown length
 * 10123 11538: contig of 1416 bp in length

11539 11638: gap of unknown length
 11639 12973: contig of 1335 bp in length
 12974 13073: gap of unknown length
 13074 14297: contig of 1224 bp in length
 14298 14397: gap of unknown length
 14398 15608: contig of 1211 bp in length
 15609 15708: gap of unknown length
 15709 17207: contig of 1499 bp in length
 17210 17307: gap of unknown length
 17308 18882: contig of 1575 bp in length
 18883 18982: gap of unknown length
 18983 20753: contig of 1771 bp in length
 20754 20853: gap of unknown length
 20854 21905: contig of 1052 bp in length
 21906 22005: gap of unknown length
 22006 23612: contig of 1607 bp in length
 23613 25358: contig of 1846 bp in length
 25359 25458: gap of unknown length
 25459 27977: contig of 2519 bp in length
 27978 28077: gap of unknown length
 28078 29889: contig of 1812 bp in length
 29890 29989: gap of unknown length
 29990 31581: contig of 1591 bp in length
 31582 33566: contig of 1886 bp in length
 33567 33666: gap of unknown length
 33667 35307: contig of 1641 bp in length
 35308 35407: gap of unknown length
 35408 37663: contig of 2356 bp in length
 37664 37863: gap of unknown length
 37864 38877: contig of 1014 bp in length
 38878 38977: gap of unknown length
 38978 41067: contig of 2090 bp in length
 41068 41667: gap of unknown length
 41668 43639: contig of 2472 bp in length
 43640 43739: gap of unknown length
 43740 46078: contig of 2339 bp in length
 46079 46178: gap of unknown length
 46179 48378: contig of 2200 bp in length
 48379 48478: gap of unknown length
 48479 50170: contig of 1692 bp in length
 50171 50270: gap of unknown length
 50271 53202: contig of 2332 bp in length
 53203 53302: gap of unknown length
 53303 56020: contig of 2718 bp in length
 56021 56120: gap of unknown length
 56121 59634: contig of 3534 bp in length
 59635 59734: gap of unknown length
 59735 61539: contig of 1805 bp in length
 61540 61639: gap of unknown length
 61640 63419: contig of 1760 bp in length
 63420 63519: gap of unknown length
 63520 66012: contig of 2493 bp in length
 66013 66112: gap of unknown length
 66113 69383: contig of 3271 bp in length
 69384 69483: gap of unknown length
 69484 72141: contig of 2658 bp in length
 72142 72241: gap of unknown length
 72242 74600: contig of 2359 bp in length
 74601 74700: gap of unknown length
 74701 78269: contig of 3569 bp in length
 78270 78369: gap of unknown length
 78370 82451: contig of 4082 bp in length
 82452 82551: gap of unknown length
 82552 86231: contig of 3740 bp in length
 86232 86331: gap of unknown length
 86332 88846: contig of 2455 bp in length
 88847 88946: gap of unknown length
 88947 92246: contig of 3300 bp in length
 92247 92347: gap of unknown length
 92348 95087: contig of 2741 bp in length
 95088 95187: gap of unknown length

	95188	98962:	contig	of	3775	bp	in	length
*	98962:	gap	of	unknown	length			
*	99063	101790:	contig	of	2728	bp	in	length
*	99063	101791:	contig	of	unknown	length		
*	101791	101890:	gap	of	unknown	length		
*	101891	104444:	contig	of	2554	bp	in	length
*	104445	104544:	gap	of	unknown	length		
*	104545	108537:	contig	of	3993	bp	in	length
*	108538	108637:	gap	of	unknown	length		
*	108638	111569:	contig	of	2932	bp	in	length
*	111570	111669:	gap	of	unknown	length		
*	111670	114896:	contig	of	3227	bp	in	length
*	114897	114996:	gap	of	unknown	length		
*	114997	118775:	contig	of	3779	bp	in	length
*	118776	118875:	gap	of	unknown	length		
*	118876	122928:	contig	of	4053	bp	in	length
*	122929	123028:	gap	of	unknown	length		
*	123029	126528:	contig	of	3500	bp	in	length
*	126529	126628:	gap	of	unknown	length		
*	126629	129982:	contig	of	3354	bp	in	length
*	129982	129982:	contig	of	3354	bp	in	length

Alignment Scores:

Alignment Scores:			
Prob. No.:	3.22E-23	Length:	195287
Score:	544.50	Matches:	136
Percent Similarity:	31.10%	Conservative:	3
Best Local Similarity:	30.43%	Mismatches:	3
Query Match:	9.69%	Indels:	305
DB:	2	Gaps:	1

US-09-830-837-6 (1-1052) x AC112476 (1-195287)

Qy	1	MetLysLeuValAsnIleTrrpLeuLeuLeuValLeuLeuCysGlyLysHis	20
Db	175393	ATGAAGCTCGTCACATCTGGCTTCTTGCTGTGGTTTGCCTGTGGGAAAAGCAT	175394
Qy	21	LeuClYasPArgLeuGluLysLysSerPheclulysLaProCysProGlyCysSerHis	40
Db	175333	CTGGGTACAGGCTGGGGAGAAAGCTTTTGAAGAAGGCCCATGCCACAGCTGTTCACCAC	175274
Qy	41	LeuThrLeuLysValGluPheSerSerThrValValGluTyr	54
Db	175273	CTGACTTTGAAGGTGGAATTCTCCTCAACTGTGTGGTAATATGTAAGGACTTTATTATCCAC	175214
Qy	54	-----	54
Db	175213	TGTTGGCCCPCCATCTCACTGTTTTCAAGCTTAGATAAATACTGGCTTCOGTTCCCTT	175154
Qy	54	-----	54
Db	175153	CTGTGTTACATTTAAATAGGCTCAGTGCATGCCATGTCGACATTAACCTTGTGTTGCTT	175094
Qy	54	-----	54
Db	175093	AAGGCGTTTGGCCTGGATTAGAGCTGACTGTACCATTGTAGCGGAGGGAACCGATGCAGG	175034
Qy	54	-----	54
Db	175033	TGCTCAGCTCAGGTTTTTGGAGGTCAGGCCTTGGGCAGACGGGCAGAGCAGGAAGTTGTC	174974
Qy	54	-----	54
Db	174973	TTGATCGGAGCTTTTGCTTCACATTCGGTAGACAGAGAAGTGCATATTCTCTGAATCCC	174914
Qy	54	-----	54
Db	174913	TCFTCCTGGPAGGAAACTACATTTCTTTACTCTAGCCGAATTTTAGTAAATATGAAGGTT	174854
Qy	54	-----	54
Db	174853	TTATCTATTTTACCTTATTATTCTTTTAAAGTTTGCAATGATTTTGGCCACATGCAG.	174794
Qy	54	-----	54
Db	174793	ACCACATGAGAGCTTGGCACTTGGCCAGCGAAGGTTGTGGGACTAGATTTCTAGTGCTGTT	174734

THIS PAGE BLANK (03/70)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 2, 2003, 01:36:29 ; Search time 3332 Seconds

(without alignments)
5113.341 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MRLVNIWLLLVLLCGKKH.....PRVKRPLMQQVHPKPTPSV 1052

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet_p2n.model -DEV=xlh
-Q/cgn2_1/USFTO.spool/US09830837/runat_23052003.181924.8071/app_query.fasta_1.1223
-DB=EST -QPM=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pcpt -THR_MAX=100 -THR_MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830837 @CGN.1.1.2475 @runat_23052003.181924.8071 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1563	27.8	927	14	BQ921443
2	1510	26.9	887	14	BQ879057
3	1470	26.2	885	9	AU140223
4	1440	25.6	1047	13	BM552158
5	1434	25.5	918	14	BQ929998
6	1418	25.2	949	13	BI905658
7	1410.5	25.1	928	14	BQ955268
8	1399.5	24.9	838	14	BQ228507
9	1380	24.6	795	13	BI870081
10	1380	24.6	870	14	BQ217956
11	1376.5	24.5	971	13	BM475146
12	1372	24.4	792	14	BQ573358
13	1349.5	24.0	944	13	BI904898
14	1349	24.0	859	14	BQ688754
15	1348	24.0	858	13	BI084430
16	1340	23.9	766	13	BI914542
17	1319	23.5	812	10	BE386747
18	1295.5	23.1	929	13	BI414585
19	1294	23.0	975	14	BQ957796
20	1293.5	23.0	859	14	BQ928302
21	1291	23.0	736	12	BG697188
22	1289	22.9	733	14	BM949214
23	1279.5	22.8	809	9	AU139053
24	1254	22.3	724	13	BI661522
25	1248	22.2	880	13	BI091766
26	1244	22.1	741	14	BM945034
27	1237	22.0	965	12	BG482068
28	1232	21.9	783	13	BI455980
29	1231	21.9	721	9	AU133912
30	1222	21.8	862	13	BI413549
31	1217	21.7	698	12	BG695170
32	1210	21.5	729	14	BM963520
33	1209	21.5	666	14	BM783543
34	1209	21.5	666	14	BM792735
35	1200.5	21.4	1010	14	BQ689579
36	1192.5	21.2	804	9	AU119413
37	1181	21.0	647	9	AL046669
38	1180	21.0	652	14	BM829953
39	1174	20.9	691	9	AU118931
40	1165	20.7	688	9	AU123483
41	1164.5	20.7	960	12	BG023875
42	1163	20.7	697	10	BE313242
43	1156	20.6	954	14	BQ645116
44	1149	20.5	835	14	BQ443546
45	1147	20.4	840	13	BI905329

ALIGNMENTS

RESULT 1
BQ921443

LOCUS

DEFINITION AGENCOURT_8930283 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6466228
5', mRNA linear EST 20-AUG-2002

ACCESSION BQ921443

VERSION BQ921443.1

KEYWORDS EST

SOURCE house mouse

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 927)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1M13990 row: d column: 05
High quality sequence stop: 697.

FEATURES
source

Location/Qualifiers
1. 927
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6466228"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
227 a 236 c 251 g 212 t 1 others

BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 3.77e-147 Length: 927
Score: 1563.00 Matches: 252
Percent Similarity: 97.07% Conservative: 6
Best Local Similarity: 95.11% Mismatches: 8
Query Match: 27.83% Indels: 1
DB: 14 Gaps: 0

US-09-830-837-6 (1-1053) x BQ921443 (1-927)

QY 577 ValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetVal 696
Db 3 GTGCTGGCGGCCCATTCACATGTTTGACGCCACACAGATATGGCAGCTTCTGCTGGTG 62
QY 697 AspSerGluGluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsn 716
Db 63 GACAGTGAGGAGAGTACTTCCCTGAGGAGATTCCTAAGCTGAGGAGGATGTGACAAAT 122
QY 717 GlyLeuSerLeuValIlePheSerAspTyrPyrAsnThrSerValMetArgIysValLys 736
Db 123 GGCTTTCTCTGTCATCTTCAGTGACTGGTACACACTTCGTATGAGAAAGTGAAG 182
QY 737 PheTyrAspGluAsnThrArgGlnTyrPheMetProAspThrGlyClyAlaAsnIlePro 756
Db 183 TTTTATGATGAACACCCAGGAGGTGGTGGATGCCACACCGGAGGAGGAGCAATCCCA 242
QY 757 AlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGly 776
Db 243 GCTCTGAATGAGCTGCTGCTGTGTGACACATGGGGTTCAGTGACGGCTATATGAAGG 302
QY 777 GluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPhe 796
Db 303 GAGTTGTCTCGCAACCATGACATGATGACTATGCTGGGGTGCAGCATCGCCCAAGTTT 362
QY 797 ProGluAspGlyValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLys 816
Db 363 CCAGAAGATGGCGTGTGATCACACAGACTTTCAGGACCAAGGATGGAGGTCTTAAA 422
QY 817 GluGluThrAlaValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGlu 836
Db 423 CAAGACAGCAGTGTGGAAATGTTCCCATTTGGGGCTTTACAGATTCATCTGAA 482
QY 837 GlyClyClyArgIleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGln 856
Db 483 GTGGAGGCGCGATGCTGCTGTATGGAGACTCCAACTGCTTGATGACAGTCACAGACAG 542

QY 857 LysAspCysPheThrPleuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrPro 876
Db 543 AAGGACTGCTTTTGGCTTCTGGATCGCTCTTCAGTACACATCTATGGGTGACCCCT 602
QY 877 ProSerLeuSerHisSerGlyAsnArgGlnArgProSerSerGlyAlaGlySerValThr 896
Db 603 CCCAGCTTCAGCCATCAGGAAACCGGACGCCACCTAGCCAGGACCGGCTTGGCCCT 662
QY 897 ProGluArgMetGluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeu 916
Db 663 CCGTGAAGATGGAAGAAACCCACCTCCATCGTACTCCCAAGTCTTGAAGCCCATGTG 722
QY 917 GlyAspProLysProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGln 936
Db 723 GGAGACCCGAAACCTCGGCCCTGCGCAGCTGTCCACATTTGTTCATGGCCCAAGCCACAG 782
QY 937 -ProLeuAsnGluThrAlaProSerAsnLeuTrpLysHisGlnLysLeuSerIleAs 956
Db 783 CCCTTTGAATGAGACGCCACCCAGTAATCTTTTGGAAACATCAGAAGCTGCTCTCCATTGA 842
QY 956 pLeuAspLysValValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSe 976
Db 843 CCTGCACAAAGTAGTGTACCCCACTTTTCGATCCCAATCGCCTCAAGTGAGAAGCTTGTG 902
QY 976 rProGlyGluSerGlyAla 982
Db 903 CCTGGAAGAGAGTGTGTC 921

RESULT 2

BQ879057
LOCUS

DEFINITION

5', mRNA sequence.

ACCESSION

BQ879057

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES
source

1. 887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6085736"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pORG7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
203 a 231 c 237 g 215 t 1 others

BASE COUNT

161 ProLeuArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis 180
 361 CCCTCGCAAGAGCAGCTCTCCCTGGGCTCTGGCTTGGCATGCTACGGAAGGCAT 420
 181 SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
 421 TCAGCAGACGGCTGCTGAGAGCCATCCGCGCCAGGTTGCCAGACACTGCAGGCAGAT 480
 201 ValLeuTrpGlnMetGlyThrGlyAlaAsnValArgValAlaValPheAspThrGly 220
 481 GTGCTCGCAGATGGATATACAGTCTTAATGTAAGATGCTGTTTGGACATGGG 540
 221 LeuSerGlyLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240
 541 CTGAGCGAGAGCATCCCACTTCAAAATGTGAGGAGAGAACCACTGGACCAACGAG 600
 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
 601 CGAACCTGGAGATGGGTGGGCATGGCACATGCTGGCAGGTGTGATAGCCAGCATG 660
 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
 661 AGGAGTCCCAAGGATTTGCTCCAGATCCAGAACATTCACATTTTCAGGGTCTTTACCAAT 720
 281 AsnGlnValSerTrpThrSerTrp-PheLeuAspAlaPheAsnTrpAlaIleLeuLys 300
 721 AATCAGGATCTTACACATCTTGGGTTTTTGGCGCGCTTAACTATGCCATTTTAAAGAA 780
 300 sIleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHis-PropheVala 320
 781 GATCGACGTGTAAACCTCAGCATTTGGCGNCCGAGATTTATGGATCATCTCCGTGGTTG 840
 320 sPlyValTrpGluLeuThrAlaAsnValIleMetValSer 334
 841 ACCANGGTGGGAATACAGCTTACATGGATCATCGGTCT 884

RESULT 4

BM552158
 LOCUS 1047 bp mRNA linear EST 20-FEB-2002
 DEFINITION AGENCOURT_6573815 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5476885
 5' mRNA sequence.

ACCESSION BM552158
 VERSION 1
 KEYWORDS GI:18789785

SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue procurement: ATCC

cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: L10M1994 row: a column: 14
 High quality sequence stop: 605.

FEATURES

Location/Qualifiers
 1..1047
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:5476885"
 /clone_lib="NIH_MGC_98"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:

ECORI: cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GCCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

BASE COUNT 247 a 265 c 305 g 230 t
 ORIGIN

Alignment Scores:

Pred. No.: 1 36e-134 Length: 1047
 Score: 1440.00 Matches: 291
 Percent Similarity: 87.17% Conservative: 29
 Best Local Similarity: 84.84% Mismatches: 15
 Query Match: 25.64% Indels: 8
 DB: 13 Gaps: 8

US-09-830-837-6 (1-1052) x BM552158 (1-1047)

QY 673 TyrPheValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThr 692
 DB 1 TACTTTGTAGAGGTCTCGGGGCCCTTCACGTGTTTGTATGCCAGTCAGTATGGCACT 60
 QY 693 LeuLeuMetValAspSerGluGluTyrPheProGluGluIleAlaLysLeuArgArg 712
 DB 61 TTGCTGATGGTGGACAGTGAAGGAGGAGTACTTCCCTGAAGAGATCGCAAGCTCCGGAGG 120
 QY 713 AspValAspAsnGlyLeuSerLeuValIlePheSerAspTrpTrpAsnThrSerValMet 732
 DB 121 GAGCTGGACAACGCCCTCTCGCTCGTCATCTTCAGTGACTGGTACAACTCTCTCTATG 180
 QY 733 ArgLysValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGlyGly 752
 DB 181 AGAAAGTGAAGTTTATGATGAACACACAGGAGTGGTGGATGCCGATACCGGAGGA 240
 QY 753 AlaAsnIleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGly 772
 DB 241 GCTACATCCACGCTCTGAATGAGCTGCTGTGTGGAACTATGGGTTCAGCATGGC 300
 QY 773 LeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTrpAlaSerGlyCysSer 792
 DB 301 CTGTATGAAGGGAGTTTCACTCCCTGGCCAACTACATGATGATATATCGTCAGGGTGACG 360
 QY 793 IleAlaLysPheProGluAspGlyValValIleThrGlnThrPheLysAspGlnGlyLeu 812
 DB 361 ATCCGGAAGTTTCCAGAAAGATGGCGTGTGATAACACAGACTTTTCAAGGACCAAGGATG 420
 QY 813 GluValLeuLysGlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyrGln 832
 DB 421 GAGGTTTAAAGCAGGAACACAGCAGTTGTGAAAACGTCCTCCATTTTGGGACTTTATCAG 480
 QY 833 IleProAlaGluGlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAspAsp 852
 DB 481 ATTCCAGCTGAGGGTGGAGCGCGGATTTGCTGATGGGACTTCCAATTTGCTGGATGAC 540
 QY 853 SerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuLeuGlnTrpSerTyr 872
 DB 541 ACTCACCCACAGAGGACTGCTTTGGCTTCTGGATGCCCTCTCCAGTACACATCGTAT 600
 QY 873 GlyValTrpProProSerLeuSerHisSerGlyAsnArgGlnArgProSerGlyAla 892
 DB 601 GGGGTGACRCGCCCTAGCTTCACTCTGGAAACGCCAGCCGCCCTCCAGTGGAGCA 660
 QY 893 GlySerValThrProGluArgMetGluGlyAsnHisLeuHisArgTyrSerLysValLeu 912
 DB 661 GGCTCAGTCATCTCCAGAGAGTGAAGGAACCACTTCTCATCGGTACTTCCAAAGGCTG 720
 QY 913 GluAlaHisLeuGlyAspProLysProArgProLeuProAlaCysPro-ArgLeuSerTr 932
 DB 721 GGAGGCCATTTGGGAGGAGCCCAACCTCGGCTCTTACACGCTGTGCCAGCTGTCTGTG 780
 QY 932 pAlaLysProGlnProLeu---AsnGluThrAlaProSerAsnLeuTrp-LysHisGlnL 951

FEATURES
source

Location/Qualifiers
1. .928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6204905"
/clone_lib="Lupski_sciatic_nerve"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dr priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCG-3' and
5'-GACTAGTTCAGATCGGAGCGGCGCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."

BASE COUNT 226 a 228 c 258 g 211 t 5 others
ORIGIN

Alignment Scores:

Pred. No.: 1.06e-131 Length: 928
Score: 1410.50 Matches: 287
Percent Similarity: 92.90% Conservative: 1
Best Local Similarity: 92.58% Mismatches: 9
Query Match: 25.11% Indels: 13
DB: 14 Gaps: 3

US-09-830-837-6 (1-1052) x BQ955268 (1-928)

QY 647 AsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgAspMetTyrGln 666
DB 2 AATGACCCCTTACCTGGAGTGGATGATCATCCACCAATTCAGGAGATATGACCAAG 61
QY 667 HisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThrCysPheAsp 686
DB 62 CATCTGAGAAGCATGGCTACTTTGTAGAGGTCTCTGGGGCCCTTCACGGTTTGTAT 121
QY 687 AlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluTyrPheProGluGlu 706
DB 122 GCACATCATATGGCACTTTGTGTATGGTGGACAGTGGAGAGTACTTCCTCTGAAGAG 181
QY 707 IleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeuValIlePheSerAspTrp 726
DB 182 ATCGCAAGCTCCGGAGGACGTGGACAACGGCCCTCTCGCTCATCTTCAGTACTGG 241
QY 727 TyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThrArgGlnTrpTrp 746
DB 242 TACAACACTCTGTTATGAGAAAGTGAAGTTTATGATGAAACACAGGAGGAGTGG 301
QY 747 MetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLeuSerValTrpAsn 766
DB 302 ATGCGCGATACCGGAGGAGCTAAATCCAGCTAAATGAGCTGCTGCTGTGTGAAC 361
QY 767 MetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyr 786
DB 362 ATGGGGTTCAGGATGGCTGTATGAAGGGAGTTCACCTGGCCCAACCATGACATGTAT 421
QY 787 TyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValIleThrGlnThr 806
DB 422 TATGGTCAAGGTGAGCATCCGGAAGTTTCAGAGATGGGCTGATACACAGACT 481
QY 807 PheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAlaValValGluAsnValPro 826
DB 482 TTCAAGGACCAAGGATGGAGGTTTAAAGCAGGAACAGCAGTGTGTTGAAACGTCGCC 541
QY 827 IleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArgIleValLeuTyrGlyAsp 846
DB 542 ATTTGGGACTTTATCAGATCCAGTGAAGGTGGAGCGCGATTGTACTGTATGGGAC 601

QY 847 SerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeu 866
DB TCCAATTGCTTGGATGACAGTCAACGACAGAGGACTGCTTTTGGCTTCTGGATGCCCTC 661
QY 867 LeuGlnTyr-ThrSerTyrGlyValThrProProSerLeuSerHisSerGlyAsnArgG 886
DB 662 CTCCAGTACATCATGATGGGTGACACCGCTAGCCTCAGTCACTCTGGGAACGCCA 721
QY 886 nArgProProSerGlyAlaGlySerValThrProGlu-ArgMetGlu-GlyAsnHisLeu 905
DB 722 GGGCCCTCCAGTGGAGGAGTCACTCACTCCAGAGAGGATGGAAAGAAACCACTCTT 781
QY 906 HisArgTyr-SerLys-ValLeuGluAlaHis----LeuGlyAspPro-LysProArg-Pr 923
DB 782 CATCGTACTCCCAAGGTTCTGGNANGCCCATTTTGGGAGACCCCAAAACCTCGNCC 841
QY 923 OleuPro-AlaCys-ProArgLeuSer---Trp-AlaLysProGlnProLeu---AsnG 940
DB 842 TCTACCAAGNCCTGTTCCAGGGTGTGCTTGGGGCCAGGCCACAGCCCTTTAAACAG 901
QY 940 uThrAlaPro 943
DB 902 AACGGGCC 911

RESULT 8

BO228507 838 bp mRNA linear EST 02-MAY-2002
LOCUS AGENCOURT_7593100 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6070621
DEFINITION 5', mRNA sequence.

ACCESSION BO228507
VERSION BO228507.1 GI:20409907
KEYWORDS EST.
SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 838)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov

Tissue Procurement: DCTD/DTF/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13355 row: d column: 14
High quality sequence stop: 606.

FEATURES

source

Location/Qualifiers
1. .838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6070621"
/clone_lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dr.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 214 a 217 c 213 g 193 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 1.15e-130 Length: 838
Score: 1399.50 Matches: 272
Percent Similarity: 83.28% Conservative: 2
Best Local Similarity: 82.67% Mismatches: 4
Query Match: 24.92% Indels: 52
DB: 14 Gaps: 2


```

Db      302 AGAGTTCCTGCTTTTGGACATCGGCTGAGCGAGAGCATCCCACTTCAAAATGTGAAG 361
QY      233 GluArgThrAsnTrpThrAsnGluArgThrLeuAspGlyLeuGlyHisGlyThrPhe 252
Db      362 GAGAGAACCACTGGACCAACGAGCGAGCGTGGACGATGGGTGGGCCATGGCACATTC 421
QY      253 ValAlaGlyValIleAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeu 272
Db      422 GTGGCAGGTGTGATAGCCAGCATGAGGAGTGCCAAAGGATTGCTCCAGATGCAGAACTT 481
QY      273 HisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAspAla 292
Db      482 CACATTTTCAGGGCTTTTACCATAATCAGGTATCTTACACATCTTTGGTTTGGACGCC 541
QY      293 PheAsnTyrAlaIleLeuLysIleAspValLeuAsnLeuSerIleGlyGlyProAsp 312
Db      542 TTCACATATGCCATTTTAAAGAGATCGACGTGTAAACCTCAGCATCGCGGCCGGAC 601
QY      313 PheMetAspHisProPheValAspLysValTrpGluLeuThrAlaAsnValIleMet 332
Db      602 TTCATGGATCATCCGTTGTTGACAAAGGTGTGGGAATTAAACAGCTAACCAATGTAATCATG 661
QY      333 ValSerAlaIleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGln 352
Db      662 GTTCTGCTATTGGCAATGACGACCTCTTTATGGCACTCTGTAATAACCCCTGCTGATCAA 721
QY      353 MetAspValIleGlyValGlyIleAspPheGluAspAsnIleAlaArgPheSerSer 372
Db      722 TGGATTGTGATGGAGTACGGCATTTGACTTTGAGATTAACATCGCCGCTTTCTTCA 781
QY      373 ArgGlyMetThr 376
Db      782 AGGGAGTGAAT 793

```

RESULT 10

```

LOCUS      BQ217956
DEFINITION AGENCOURT_7546723 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6025653
5', mRNA sequence.
ACCESSION  BQ217956
VERSION     BQ217956.1 GI:20399356
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 870)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM13238 row: b column: 22
High quality sequence stop: 709.
Location/Qualifiers
1..870
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6025653"
/clone_lib="NIH_MGC_70"
/tissue_type="epitheloid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."

```

FEATURES

source

```

BASE COUNT  228 a  196 c  237 g  207 t  2 others
ORIGIN
Alignment Scores:
Pred. No.:    1.16e-128      Length:      870
Score:        1380.00       Matches:    265
Percent Similarity: 95.39%   Conservative: 4
Best Local Similarity: 93.97% Mismatches:    8
Query Match:  24.57%       Indels:     5
Db:           14           Gaps:      1
US-09-830-837-6 (1-1052) x BQ217956 (1-870)
QY      564 ILeSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisVal 583
Db      2 ATCTCCATTCCTGTGACCAAGAAGCGCTCTCTGGGAAGGCATTGCTCAGGGCCATGTC 61
QY      584 MetIleThrValIleAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThrSer 603
Db      62 ATGATCACTGTGGCTTCCCAGCAGACAGACAGATCAAAAATGGTGCAGAACAGACTTCA 121
QY      604 ThrValLysLeuProIleLysValLysIleIleProThrProProArgSerLysArgVal 623
Db      122 ACAGTAAAGCTCCCAATTAAGTGAAGATTAATTCCTACTCCCCCGGAGCAAGACAGACTT 181
QY      624 LeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArgAspAsnLeu 643
Db      182 CTCTGGGATCAGTACCACCAACCTCCGCTATCCACCTGGCTATTTCCACAGGATAATTTA 241
QY      644 ArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgAsp 663
Db      242 AGGATGAAGAATGACCCCTTTAGACTGGAATGGTGGATCAGATCCATCCACACCAATTTTC 301
QY      664 MetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThr 683
Db      302 ATGTACAGCATCTGAGAAGCATGGGCTACTTTGTAGAGGTCTCTCGGGGCCCTTCACG 361
QY      684 CysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGlyThrPhe 703
Db      362 TGTGTTGATGCCAGTCAGTATGCACTTGCTCATGTGTGACAGTGGAGGAGTACTTTC 421
QY      704 ProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeuValIlePhe 723
Db      422 CCTGAAGAGATCCGCAAGCTCCGGAGGAGCGTGGACAAACGGCTCTCGCTCGTCATCTTC 481
QY      724 SerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThrArg 743
Db      482 AGTGACTGTACAAACACTTCTGTTATGAGAAAAGTGAAGTTTATGATGAAAACACAGG 541
QY      744 GlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLeuSer 763
Db      542 CAGTGTGGATGCGGATACCGGAGGAGCTAACATCCAGCTCTGAATGAGCTGTCTGTCT 601
QY      764 ValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHis 783
Db      602 GTGTGAACATGGGGTTACGCGATGGCTGTATGAAGGGGAGTTACACCTGGGCCAACCAT 661
QY      784 AspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValIle 803
Db      662 GACATGTATTATGCGTCAGGGTCAGCATCGCGAGTTTCCAGAAGATGGCGTCGTGATA 721
QY      804 ThrGlnThrPheLysAspGln-GlyLeuGluValLeuLys-----GlnGluThrAlaVal 821
Db      722 ACACAGACTTTTCAAGGACCAAGGATTTGGAGTTTANAGCAGGAGGAACACAGCCAGTTGT 781
QY      821 lValGluAsnValProIleLeuGly--LeuTyrGlnIleProAlaGluGlyGlyArg 840
Db      782 TTGAAAAAAGCGTCCCATTTTGGGAACCTTTATCAGATTCCACGCTTGGAGGTGGGAAG 841
RESULT 11
BM475146
LOCUS      BM475146
DEFINITION AGENCOURT_6477357 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5559245

```

```

5', mRNA sequence.
BM475146
BM475146.1 GI:18524188
EST.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 971)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rga@bbs-rcmail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM12283 row: i column: 06
High quality sequence stop: 650.
Location/Qualifiers
1. 971
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5559245"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ (small intestine; Vector: pCMV-SPORT6;
Site_1: Noti; Site_2: Sali; Cloned unidirectionally;
oligo-dt primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 234 a 240 c 276 g 220 t 1 others
ORIGIN

Alignment Scores:
Pred No.: 3, 22e-128 Length: 971
Score: 1376.50 Matches: 287
Percent Similarity: 89.78% Conservative: 3
Best Local Similarity: 88.85% Mismatches: 17
Query Match: 24.51% Indels: 16
Dbs: 13 Gaps: 6

US-09-830-837-6 (1-1052) x BM475146 (1-971)

QY 664 MetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThr 683
DB 5 ATGTACCAAGCATCTGAGAACGATGGCTACTTTGTAGAGGTCTCTCGGGGCCCTTCACG 64
QY 684 CysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGluTyrPhe 703
DB 65 TGTGTTGATGCCAGTCAGTATGGCACTTTCCTGATGGTGGACAGTGAGGAGTACTTC 124
QY 704 ProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeuValIlePhe 723
DB 125 CTGAAGAGATGCCAAGCTCGGAGGGAGCTGGACACGGCCCTCTCGCTGCTCATCTTC 184
QY 724 SerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThrArg 743
DB 185 AGTGACTGGTACACACTTCGTGTATGAGAAAGTGAAGTATGATGAAACACCAAGG 244
QY 744 GlnTrpTrpMetProAspThrGlyClyAlaAsnIleProAlaLeuAsnGluLeuLeuSer 763
DB 245 CAGTGGTGGATGCCGATACCGAGGAGCTACATCCAGCTCTGAATGAGCTGTCT 304
QY 764 ValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHis 783
DB 305 GTGTGAACATGGGGTTCAGCGATGGCCCTGTATGAAGGGGAGTTCACCCCTGGCCCAACCAT 364

```

```

784 AspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValIle 803
DB 365 GACATGTATTATCGTCAGGGTCAGCATCGGAAGTTTCCAGAAGATCGCGCTCGTGATA 424
QY 804 ThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAlaValIleGlu 823
DB 425 ACACAGACTTTTCAAGGACCAAGATTGGAGCTTTTAAAGCAGGAAACAGCAGTGTGTGAA 484
QY 824 AsnValProIleLeuGlyLeuTyrGlnIleProAlaGluClyGlyValIleValIle 843
DB 485 AACGTCCTCCATTTTGGGACTTTATCAGATTCACGCTGAGGGTGGAGCGGATGCTACTG 544
QY 844 TyrGlyAspSerAsnGlyCysLeuAspSerHisArgGlnLysAspCysPheTrpLeuLeu 863
DB 545 TATGGGAGCTCCAAATGCTTGGATGACAGTCCGACGAGGAGCTGCTTTTGGCTCTG 604
QY 864 AspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProSerLeuSerHisSerGly 883
DB 605 GATGCCCTCTCCAGTACACATCGTATGGGTGACCGCTAGCTAGCTCAGTCACTCTGGG 664
QY 884 AsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMetGlu-GlyAs 903
DB 665 AACCCGCCAGGCGCTCCAGTGGAGCAGCTCAGTCACTCCAGAGAGTGAAGAGAA 724
QY 903 nHisLeuHisArgTyr-SerLys-ValLeuGluAla-HisLeu-GlyAspProLysPro-- 921
DB 725 ACATCTTCATCGGTACTCCAAAGGTTCTGTGAGGCCCTTGTGGGAGACCCCAAAACCTT 784
QY 922 -ArgProLeuProAla-CysProArg-LeuSerTrp-AlaLysProGlnProLeu--As 939
DB 785 CGGGCTCTACCAAGCGGCTCCAGCGCTTGTCTTGGGCGCAAGCCCGCTTTTAAAC 844
QY 939 nGluThr---AlaProSerAsnLeuTrp---LysHisGlnLys-----LeuLe 953
DB 845 CGAGACGGCGGCCCAAGTAACTTTTGGAAACATACAGAAAGCTACTCTCCCATGA 904
QY 953 uSerIleAspLeuAspLysValValLeuProAsnPhe---ArgSerAsnArgPro 970
DB 905 ACCNFGGACACAGGGGGGTGTGTTAACCAACCTTTTCGGAATCGAATCGGCC 959

RESULT 12
PQ573358 792 bp mRNA linear EST 19-JUN-2002
LOCUS UI-M-FD0-byg-b-18-0-UI.r1 NIH_BMAP_F00 Mus musculus cDNA clone
DEFINITION IMAGE:5717681.5', mRNA sequence.
ACCESSION PQ573358
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 792)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rga@bbs-rcmail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyx-5.
Location/Qualifiers
1. 792
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"

```

```

FEATURES
source
Location/Qualifiers
1. 792
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"

```



```

/clone="IMAGE:5717681"
/clone_lib="NIH_BMAP_FDO"
/tissue_type="whole brain"
/dev stage="embryo 12.5 dpc"
/lab host="DHI0B (T1 phase resistant)"
/note="Organ: brain; Vector: pfx-Asc; Site:1: EcoR I;
Site:2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pfx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TCAGAGAGCC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 193 a 208 c 215 g 174 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 6.33e-128 Length: 792
Score: 1372.00 Matches: 251
Percent Similarity: 98.11% Conservative: 8
Best Local Similarity: 95.08% Mismatches: 5
Query Match: 24.43% Indels: 0
DB: 14 Gaps: 0

US-09-830-837-6 (1-1052) x BQ573358 (1-792)

Qy 514 ValAsnValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAsp 533
Db 1 GTTAATGTCACCATCCATCGATGGATGGCGTCACAGGAAGATGTGGATAAGCCCTGAG 60

Qy 534 TrpGlnProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSer 553
Db 61 TGGCGACCCATTATACACAGATGGAGACAACTTGAAGTGGCTTCTCTACTCTCTCA 120

Qy 554 ValLeuTrpProTyrSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaAla 573
Db 121 GTGTGTGGCCCGTGGTCCAGTTACCTTGGCATCTCCATTTCTGTGACCAAGAGCCAGCT 180

Qy 574 SerTrpGluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThr 593
Db 181 TCCTGGGAAGCATCGCTCAGGCCACATCATGATCACATGGGTCCCGCCAGCAGACA 240

Qy 594 GluSerLysAsnGlyAlaGluGlnThrSerThrValLysLeuProIleLysValLysIle 613
Db 241 GAGTTACACAGTGGTGGGAGCACACTTCCACCGTGAAGCTGCCATCAAGTGAAGATC 300

Qy 614 IleProThrProProArgSerLysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyr 633
Db 301 ATTCCCAACCCCTCTCCGAGCAAGAGAGTCTCTCGGACCAAGTACCAACACCTCCGCTAC 360

Qy 634 ProGlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAspTyrAsn 653
Db 361 CCACCTGGCTACTCCCGAGGACAACCTTCGGATGAAGATGACCCCTTAGACTGGAAT 420

Qy 654 GlyAspHisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTyr 673
Db 421 GGGGACACCTCCACACCACTTCAGGACATGTACCAGCATGTGGCAGCATGGGCTAC 480

Qy 674 PheValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeu 693
Db 481 TTCGTGGAGGTGTGGCGGCCCATTCACATGTTTTCAGCCACACACAGTAGTGCACCTTG 540

Qy 694 LeuMetValAspSerGluGluGluTyrPheProGluGluIleAlaLysLeuArgArgAsp 713
Db 541 CTGCTGTGGACAGTAGGGAAGAGTACTTCCCTCGAGAGATTGCTAAGCTGAGAGGGAT 600

714 ValAspAsnGlyLeuSerLeuValIlePheSerAspTyrPyrAsnThrSerValMetArg 733
Db 601 GTGGACAATGGCCCTTTCCCTCGTCATCTTCAGTGACCTGGTACACACTTCTGTTATGAGA 660

734 LysValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGlyGlyAla 753
Db 661 AAGTGAAGTTTATGATGAAACACACAGGAGCTGGTGGATGCCAGACACCGGAGAGCG 720

754 AsnIleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGlyLeu 773
Db 721 AACATCCAGCTCGAATGAGCTGCTGTGTGTGNGACATGTTGGGGTTCAGTGACGGNCTA 780

774 TyrGluGlyGlu 777
Db 781 TATGAAGGGGAG 792

RESULT 13
BI904898
LOCUS
DEFINITION 603168937F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5257041 5',
mRNA sequence.
ACCESSION BI904898
VERSION BI904898.1 GI:16167312
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 944)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-femail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1649 row: i column: 10
High quality sequence start: 36
High quality sequence stop: 926.
FEATURES
Location/Qualifiers
source
1..944
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5257041"
/clone_lib="NCI_CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DHI0B (phage-resistant)"
/note="Organ: lung; Vector: pT73b-Pac (Pharmacia) with a
modified polylinker; Site:1: NotI; Site:2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGGGAGGGCGGCCCTCGTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 228 a 229 c 259 g 227 t
ORIGIN

Alignment Scores:
Pred. No.: 1.66e-125 Length: 944
Score: 1349.50 Matches: 273
Percent Similarity: 91.83% Conservative: 8
Best Local Similarity: 89.22% Mismatches: 21
Query Match: 24.03% Indels: 9

```


QY 994 rAsnGlnValGlyGlnThrIleProValPheAlaPheLeuGlyAlaMetValValle 1014
 Db 783 CAC-CAGAGGTGGGCGCAGCAATCCCTGCTTTC-TTCTGTGAGCCATGGTGGCT 840

QY 1014 uAlaPheValVal 1019
 Db 841 GAGCTTCTTGGTGGA 856

RESULT 16
 BI914542 766 bp mRNA linear EST 16-OCT-2001
 LOCUS 603180473F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5244570 5',
 DEFINITION mRNA sequence.

ACCESSION BI914542
 VERSION BI914542.1 GI:16198939
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 766)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M1617 row: a column: 19
 High quality sequence stop: 738.
 Location/Qualifiers
 1..766
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5244570"
 /clone_lib="NIH_MGC_121"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI;
 Site 2: EcoRV (destroyed); RNA source anonymous pool of 3
 fetal brains, female age 20 weeks, female age 24 weeks,
 and male age 26 weeks. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.7 kb, insert size range
 0.7-3.5 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 017. Note:
 this is a NIH-MGC Library."

BASE COUNT 193 a 199 c 187 g 186 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,04e-124 Length: 766
 Score: 1340.00 Matches: 250
 Percent Similarity: 99.21% Conservative: 0
 Best Local Similarity: 99.21% Mismatches: 2
 Query Match: 23.86% Indels: 1
 DB: 13 Gaps: 0

US-09-830-837-6 (1-1052) x BI914542 (1-766)

QY 483 SerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSer 502
 Db 10 AGTTTGAGCCCGCAGTACATAGATCTGACTGAGTGTCCCTACATGTGGCCCTACTGCTCC 69

QY 503 GlnProIleTyrGlyMetProThrValValAsnValThrIleLeuAsnGlyMet 522
 Db 70 CAGCCCTACTACTATGAGGAGTCCGACAGTGTAAATGTCAACCTCTCCACAGGCATG 129

QY 523 GlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrIleuProGlnAsnGly 542
 Db 130 GGAGTCACAGAAAGAAATGTAGATAAGCTGACAGCGAGCCATTTTGCACAGACCGA 189

QY 543 AspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTrpSerGlyTyrIleu 562
 Db 190 GACAACATGGAAGTTGCTTCTCTACTCTCTGGTCTTATGGCTTGGTGGGCTACCTG 249

QY 563 AlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHis 582
 Db 250 GCATCTCCATTTCTGTGACCAAGAGCGCTTCTGGGAAGGCATTTCTCAGGCGCAT 309

QY 583 ValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThr 602
 Db 310 GTCATGATCACTGTGGCTTCCCGACAGACAGAGTCAAAAATGTCAGAACAGACT 369

QY 603 SerThrValLysLeuProIleLysValLysIleIleProThrProProArgSerLysArg 622
 Db 370 TCAACAGTAAAGCTCCCAATTAAGGTGAAGATAATTCCTACTCCCGCGCAAGAGA 429

QY 623 ValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArgAspAsn 642
 Db 430 GTTCTCTGGGATCAGTACACACACCTCCGCTATCCACCTGGCTATTTCCCGAGGATAAT 489

QY 643 LeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArg 662
 Db 490 TTAGGTGGAAGATGACCTTTAGACTGGATGGTATCATCATCCACACCATTTTCAGG 549

QY 663 AspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPhe 682
 Db 550 GATATGTACCAAGCATCTGAGAGCATGGCTACTTTGAGAGGTCTCGGGGCCCTTC 609

QY 683 ThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluTyr 702
 Db 610 ACGTGTNTGTAGTCCAGTCAGTATGCACCTTGTGTGTGGTGGACAGTGGAGGAGTAC 669

QY 703 PheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeuValle 722
 Db 670 TTCCCTGGAAGAGATCGCCAAAGCTCCGGAGGACGTGGACACACGCTC-TCGCTGCTATC 728

QY 723 PheSerAspTrpTyrAsnThrSerValMetArgLys 734
 Db 729 TTCAGTGTAGTGTACAAACACTTCTGTATGAGAAAA 764

RESULT 17
 BE386747 812 bp mRNA linear EST 21-JUL-2000
 LOCUS 601273046F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3614338 5',
 DEFINITION mRNA sequence.

ACCESSION BE386747
 VERSION BE386747.1 GI:9332112
 KEYWORDS EST
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 812)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
 Plate: L14M276 row: k column: 11
 High quality sequence start: 26
 High quality sequence stop: 812.
 Location/Qualifiers
 1..812

FEATURES
 SOURCE

Db 62 ATGTCACCTATGCTGGAGTGGGGTTCCGGTGTGAAAGGGGCGCTCCGCTCACTC 121
QY 411 SerGlyThrSerValAlaSerProValValAlaGlyAlaValThrLeuLeuValSerThr 430
Db 122 TCAGGAACAGTGTCTCCCTCCAGTGTGGTGGGCGCTCACTCTGTAGTAAGCACA 181
QY 431 ValGlnLysArgGluLeuValAsnProAlaSerMetLysGlnAlaLeuLeuAlaSerAla 450
Db 182 GTACAGAGCGGAGCTGGTGAATCTCCAGTGTGAACCAAGCTTTGATAGGCTCAGCC 241
QY 451 ArgArgLeuProGlyValAsnMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArg 470
Db 242 CGGAGACTCTCTGGGTCAACAAGTTCGACCAAGGTCATGGCAAGTGTGATCTCTGCGCA 301
QY 471 AlaTyrGlnIleLeuAsnSerTyrLysProGlnAlaSerLeuSerProSerTyrIleAsp 490
Db 302 GCTTATCAGATCTCTCCAGCTATTAACCGCAGGAGCTGAGTCTAGCTACATCGAC 361
QY 491 LeuThrGluCysProTyrMetTrpProTyrCysSerGlnProIleTyrTyrGlyGlyMet 510
Db 362 CTGACTGAGTCTCCATGATGGCCCTACTGCTCCAGCCTTACTTACTATGAGGAATG 421
QY 511 ProThrValAlaValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAsp 530
Db 422 CCAACAATCGTTAATGTACCATCTCAATGGCATGGCGTCCACAGGAAGAT-GTGGAT 480
QY 531 LysProAspTrpGlnProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSer 550
Db 481 AAGCCTGAGTGGGACCCATTTTACCACAGAATGGAGACACATTAAGTGGCTTCTCC 540
QY 551 TyrSerSerValLeuTrpProTyrSerGlyTyrLeuAlaIleSerIleSerValThrLys 570
Db 541 TACTCTCAGTGTGTGGCCCTGGTCAAGTTACCTTGTCCATCTCCATTTCTGTGACCAAG 600
QY 571 LysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerPro 590
Db 601 AAGCGAGCTCTCTGGGAGGCGATCGCTCAGGGCCACATCATCATCAGTGGGCTC-CCA 659
QY 591 AlaGluThrGluSer-LysAsnGlyValGluGlnThrSerThrValLysLeuProIleGly 610
Db 660 GCAGAGACAGATTACACAGATGGTGGCGAGCACATCTCCACCTGAGAGTGGCCATCA 719
QY 610 sValLysIleIleProThrProArgSerLysArgValLeuTrpAsp-GlnTyrHisA 630
Db 720 GGTGAAGATCATCTCCAC-CCTCCTCGGAGCAAGAGAGTCTCTGGGACCATGTACCACA 778
QY 630 snLeuArgTyrProProGlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProL 650
Db 779 AC---TCCGTACGACCTGTGTACTTCCCGAGGACAACTTGGGATGAAGATGACCTTT 835
QY 650 euAsp-TripAsnGlyAspHisIleHisThrAsnPheArgAspMetTyrGlnHis-LeuAr 669
Db 836 TAGACTTGCATTTGGGAGCAGGTCAAGAGCCAAATTCAGGGAATGTACCAGCATCTTGG 895
QY 669 gSerMetGlyTyrPheValGluValLeuGly 679
Db 896 CAGCATGGGGCTACTCGGAGGGTGTCTGGGG 926

RESULT 19
LOCUS B0957796
DEFINITION AGENCOURT_8772898 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6381912
5', mRNA sequence.
ACCESSION B0957796
VERSION B0957796.1 GI:22373274
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 975)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue procurement: DCTD/Drp
cDNA Library Preparation: Rubin Laboratory
cDNA Sequencing by: Agencourt Bioscience Corporation (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: L10K2572 row: 0 column: 01
High quality sequence stop: 535.
Location/Qualifiers
1..975
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6381912"
/clone_lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library."
BASE COUNT 218 a 280 c 275 g 201 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 7,41e-120 Length: 975
Score: 1294.00 Matches: 252
Percent Similarity: 93.77% Conservative: 4
Best Local Similarity: 92.31% Mismatches: 12
Query Match: 23.04% Indels: 5
DB: 14 Gaps: 2

US-09-830-837-6 (1-1052) x BQ957796 (1-975)
QY 783 HisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProLysAspGlyValVal 802
Db 2 CATGACATGATATTATGCGTCAGGTGCAGCATCCGGAAGTTTCCAGAGATGCGCTGTG 61
QY 803 IleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAlaValVal 822
Db 62 ATAACACAGACTTTCAAGGACCAAGGATTTGAGGTTTAAAGCAGGAAACAGCAGTGT 121
QY 823 GluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArgIleVal 842
Db 122 GAAACAGTCCCATTTTGGGACTTTATCAGATTCCAGCTGAGGGTGGAGCGCGATTGTA 181
QY 843 LeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTrpLeu 862
Db 182 CTGTATGGGACTCCAATTGCTTGGATGACAGTCAACCCACAGAAAGACTGCTTTTGGCTT 241
QY 863 LeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProSerLeuSerHisSer 882
Db 242 CTGGATGCCCTCTCCAGTACATCGTATGGGTGACACCGCTAGCCTCAGTCACTCT 301
QY 883 GlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProLysArgMetGluGly 902
Db 302 GGGACACCGCCAGCCCTCCAGTGGAGCAGCTCAGTCACTCCAGAGAGGATGAAGGA 361
QY 903 AsnHisLeuHisArgTyrSerLysValLeuLuhHisLeuGlyAspProLysProArg 922
Db 362 ACCATCTTCATCGGTACTCCAAAGGTCTTGGAGCCCTTTGGAGAGCCCAACCTCGG 421
QY 923 ProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGluThrAla 942
Db 422 CCTTACCAGCTGTCCACGCTGTCTTGGGCCAAGCCACAGCCTTTTAAACGAGACGGCG 481

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10699 row: h column: 06
High quality sequence start: 4
High quality sequence end: 676.
FEATURES
source
1. 736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4803509"
/lab_host="NCI_CGAP_Skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 195 a 168 c 203 g 170 t
ORIGIN
Alignment Scores:
Pred. No.: 8.89e-120 Length: 736
Score: 1291.00 Matches: 241
Percent Similarity: 99.18% Conservative: 2
Best Local Similarity: 98.37% Mismatches: 2
Query Match: 22.98% Indels: 1
DB: 12 Gaps: 0
US-09-830-837-6 (1-1052) x BG697188 (1-736)
QY 596 LysAsnGlyAlaGluGlnThrSerThyValLeuProIleValLysIlePro 615
Db 3 AAAATGGTGCAGAACAGACTTACAGTAAGCTCCCATTAAGTGAAGATAATTCCT 62
QY 616 ThrProArgSerLysArgValLeuTrpAspGlnThrHisAsnLeuArgTyrProPro 635
Db 63 ACTCCCGCCGAGCAGAGAGTCTCTGGGATCAGTACCACACCTCCGATCCACCT 122
QY 636 GlyTyrProArgAspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAsp 655
Db 123 GGCTATTTCCCGAGGATTAATTTAGGATGAGATGACCTTTAGCTGGATGGTAT 182
QY 656 HisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheVal 675
Db 183 CATATCCACCAATTTCCAGGATATGATCCAGCATCTGAGAAGCATGGCTACTTTGTA 242
QY 676 GluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMet 695
Db 243 GAGGTCTCCGGGCCCCCTTACGTGTTTGTATGCCAGTCAGTATGGCATTCTGATG 302
QY 696 ValAspSerGluGluGluTyrPheProGluGluIleAlaLysLeuArgAspValAsp 715
Db 303 GTGGACAGTGAGGAGGATCTTCTCCGAAGAGATGCCAAGCTCCGAGGAGCTGGAC 362
QY 716 AsnGlyLeuSerLeuValIlePheSerAspTrpTrpAsnThrSerValMetArgLysVal 735
Db 363 AACGGCTCTCCCTCGTCATCTTCACTGATGCTGATGATGATGATGATGATGATG 735
QY 736 LysPheThrAspGluAsnThrArgGlnTrpMetProAspThrGlyClyAlaAsnIle 755

423 AAGTTTATGATCAAAACACAGGAGGAGTGGTGGATCCGATACCGAGGAGCTAACATC 482
QY 756 ProAlaLeuAsnGluLeuLeuSerValTrpAsnMetClyPheSerAspGlyLeuTyrGlu 775
Db 483 CCAGCTCTGATGAGTCTGCTGCTGTGTGGAAACATGGGGTTCAGCGATGGCTCTATGAA 542
QY 776 GlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerClyCysSerIleAlaLys 795
Db 543 GGGGAGTTTCACTCCCTGGCCCAACCATGACATGATATATGCTGAGGTGGAGATCCGAG 602
QY 796 PheProGluAspGlyValValIleThrGlnThrPheLysAspGlnClyLeuGluValLeu 815
Db 603 TTTCCAGAAAGATGCGCTGCTGATAACACAGACTTTCAAGGACCAAGGACTGGAGTTGTA 662
QY 816 LysGlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAla 835
Db 663 AAGCAGGACACAGCAGCTGTTGAAACAGTCCCGGACTTTGAGACTTTATCATGATCCGCT- 721
QY 836 GluGlyGlyGlyArg 840
Db 722 GAGGTGGAGCGCG 736

RESULT 22
BM949214 733 bp mRNA linear EST 14-MAR-2002
LOCUS UI-M-EG0p-buy-e-17-0-UI-r1 NIH-BMAP_EG0p Mus musculus cDNA clone
DEFINITION IMAGE:5687416 5', mRNA sequence.
ACCESSION BM949214
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: Dr. James Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. 733
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5687416"
/clone_lib="NIH-BMAP_EG0p"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: brain; Vector: pYX-5; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-5 vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCACGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project

(BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 176 a 173 c 206 g 174 t 4 others
ORIGIN

Alignment Scores:

Pred. No.: 1,41e-119 Length: 733
Score: 1289.00 Matches: 234
Percent Similarity: 97.95% Conservative: 5
Best Local Similarity: 95.90% Mismatches: 5
Query Match: 22.95% Indels: 0
DB: 14 Gaps: 0

US-09-830-837-6 (1-1052) x BM949214 (1-733)

QY 632 ArgTyrProGluGlyThrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAsp 651
Db 1 CGCTACCCACCTGGGTACTTCCCGAGGACAACTTGGGATGAAGAATGACCTTTAGAC 60
QY 652 TrpAsnGlyAspHisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMet 671
Db 61 TGAATGGCGACACGCTCCACACCACTTCAGGACATGTACAGCATCTGCGACAGC 120
QY 672 GlyTyrPheValGluValLeuGlyAlaPropheThrCysPheAspAlaSerGlnTyrGly 691
Db 121 GGCTACTTCTGGAGGTGCTCGCGCCCATTCACATGTTTGACGCCACACAGATGCG 180
QY 692 ThrLeuLeuMetValAspSerGluGluGluTyrPheProGluGluIleAlaLysLeuArg 711
Db 181 ACTTTGCTGCTGGTGACAGTGAAGAGTAGTACTTCCCTGAGGAGATTCTTAAGCTGAG 240
QY 712 ArgAspValAsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThrSerVal 731
Db 241 AGGGATGGGCAATGGCCCTTCCCTCGTCATCTTCAGTACTGGTACACACTTCTGT 300
QY 732 MetArgLysValLysPheTyrAspGluAsnThrArgGlnTrpTrpMetProAspThrGly 751
Db 301 ATGAGAAAGTGAAGTTTATGATGAAACACACGAGCAGTGGTGGATGCCACACCGGA 360
QY 752 GlyAlaAsnIleProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAsp 771
Db 361 GGAGCGAATCCCACTGTAATGAGCTGCTGTGTGTGGAACATGGGGTTCAGTGAC 420
QY 772 GlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCys 791
Db 421 GGCCATATGAGGGGAGTTTCTCTGGCANACCATGACATGACTATCGCTGGGGTGC 480
QY 792 SerIleAlaLysPheProGluAspGlyValValIleThrGlnThrPheLysAspGlnGly 811
Db 481 AGCATCGCCAACTTCCAGAAGATGGCGTGTGATCACACAGACTTTCANGGACCAANGA 540
QY 812 LeuGluValLeuLysGlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyr 831
Db 541 TTGGAGGTCTTAAACAAGACAGACAGTGTGGAAAATGTTCCTATTTGGGGCTTTAT 600
QY 832 GlnIleProAlaGluGlyGlyArgIleValLeuTyrGlyAspSerAsnGlyLeuAsp 851
Db 601 CAGATCCATCTGAGGTGGAGCGCGGATCGTGTGTATGGAGACTCCCAACTGCTTGGAT 660
QY 852 AspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuLeuGlnTyrThrSer 871
Db 661 GACAGTCAAGACAGAGGAGTGTCTTGGCTTCTGGATCGCTCTTTCAGTACACATCC 720
QY 872 TyrGlyValThr 875
Db 721 TATGGCGTGACC 732

RESULT 23

AUI39053

LOCUS

DEFINITION AUI39053 809 bp mRNA linear EST 02-AUG-2002
AUI39053 PLACE1 Homo sapiens cDNA clone PLACE1009836 5', mRNA
sequence.

ACCESSION

AUI39053 GI:11000574

VERSION

EST.

KEYWORDS

human.

SOURCE

Homo sapiens

ORGANISM

human.

REFERENCE

1 (bases 1 to 809)

AUTHORS

Ota T., Nishikawa T., Suzuki Y., Ishii S., Saito K., Kawai Y., Yamamoto J., Wakamatsu A., Nakamura Y., Nagai R., Sugano S. and Isogai T.

TITLE

HRI human cDNA project

JOURNAL

Unpublished (2000)

COMMENT

Contact: Takao Isogai

Genomics Laboratory

Helix Research Institute

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3975

Fax: 81-438-52-3986

Email: genomics@hri.co.jp

HRI human cDNA project: 5' - 3' - end one pass sequencing: Helix

Research Institute; cDNA library construction: Department of

Virology, Institute of Medical Science, University of Tokyo, and

Helix Research Institute.

FEATURES

Location/Qualifiers

1..809

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="PLACE1"

/tissue_type="placenta"

/note="Vector: pME18SFL3"

211 a 194 c 213 g 185 t 6 others

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 1.55e-118 Length: 809

Score: 1279.50 Matches: 248

Percent Similarity: 93.28% Conservative: 2

Best Local Similarity: 92.54% Mismatches: 12

Query Match: 22.78% Indels: 6

DB: 9 Gaps: 1

US-09-830-837-6 (1-1052) x AUI39053 (1-809)

QY 536 ProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeu 555

Db 3 CCTATTTCACAGAACGAGACAACTTGAAGTTGCCTTCTCTACTCTCGGTCTTA 62

QY 556 TrpProTyrSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaLaSerTyr 575

Db 63 TGGCCTTGGTGGGCTACCTGGCCATCTCCATTTCTGTACCAGAAAGCGGCTCTCG 122

QY 576 GluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThrGluSer 595

Db 123 GAAGGCATTGCTCAGGGCCATGTCATGATGATGCTGTGGTCTCCACAGACAGACAGATCA 182

QY 596 LysAsnGlyAlaGluGlnThrSerThrValLysLeuProIleLysValLysIleIlePro 615

Db 183 AAAAAATGGTGCAACAGACTTCACAGATAAGTCCCTCCATTAAGTGAAGATAATTCCT 242

QY 616 ThrProProArgSerLysArgValLeuTyrAspGlnTyrHisAsnLeuArgTyrProPro 635

Db 243 ACTCCCCCGGAGCAAGAGAGTTCTCTGGGATCAGTACACACACCTCCGCTATCCACCT 302

QY 636 GlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAsp 655

Db 303 GGCTATTTCCCGAGGATTAATTAAGATGAAGATGACCTTTAGACTTGAATGCTGAT 362

QY 656 HisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheVal 675

Db 363 CACATCCACCAATTTCCAGGATATGTACCAGCATCTGAGAAGCATCTGGCTACTTTGTA 422

QY 676 GluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMet 695
 Db 423 GAGGTCCTGGGGCCCCCTTCACGGTTTTCATGCCAGTCAGTATGGCACTTTCCTGATG 482
 QY 696 ValAspSerGluGluTyrPheProGluGluIleAlaLysLeuArgAspValAsp 715
 Db 483 GTGGACAGTGGGAGGAGTACTTCCTCAAGAGATGCCAAGCTCCGGAGGGACGTGGAC 542
 QY 716 AsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThrSerValMetArgLysVal 735
 Db 543 AACGGNCTCTCGCTCGTCTCATCTTCAGTGACTGGTACACACTTCTGTATGACAAAAGTG 602
 QY 736 LysPheTyrAspGluAsnThrArgGlnTyrTrpMetProAspThrGly--GlyAlaAsnI 755
 Db 603 AAGTTTATGATGAAACACAAAGGAGTGGTGGATCCGGATACCGGGAGGAGCTAACCA 662
 QY 755 IeProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrG 775
 Db 663 TCCAGCTCTGATGAANTGCTGCTGTGTGGNACATGGGGGTCAACGATGGCTGTATG 722
 QY 775 Iu-GlyGluPheThrLeuAlaAsnHis-AspMetTyrTyrAlaSer-GlyCysSerIleA 794
 Db 723 AAAGGGGAGTCAACCTGGGCAACCATGGACATGATGATTANGGCTCAAGGTGCAACCAAC 782
 QY 794 Ialys---PhePro 797
 Db 783 CCCGAANTTTCCA 796

RESULT 24

LOCUS BI661522 724 bp mRNA linear EST 12-SEP-2001
 DEFINITION 60305833F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5351337 5',
 mRNA sequence.

ACCESSION BI661522

VERSION BI661522.1 GI:15575758

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE NIH-MGC http://mgi.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth

Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM1893 row: b column: 10

High quality sequence stop: 723.

Location/Qualifiers

1..724

/organism="Mus musculus"

/strain="NMRI"

/db_xref="taxon:10090"

/clone_image="5351337"

/clone_lib="NCI_CGAP_Mam4"

/tissue_type="tumor, gross tissue"

/dev_stage="5 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; SalI;

Site:2; NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Priscilla Furth,

NIH Reference for transgenic model: Li et al., Cell Growth

and differentiation 7, 3-11 (1996)."

175 a 212 c 181 g 156 t

BASE COUNT

ORIGIN

Alignment Scores:
 Pred. No.: 4.84e-116 Length: 724
 Score: 1254.00 Matches: 228
 Percent Similarity: 98.34% Conservatives: 9
 Best Local Similarity: 94.61% Mismatches: 3
 Query Match: 22.33% Indels: 1
 DB: 13 Gaps: 0

US-09-830-837-6 (1-1052) x BI661522 (1-724)

QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480
 Db 2 CAAGTCCATGCAAGTTGGATCTGTCGAGCTTATCAGATCTCAGCAGCTATAAACCG 61
 QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
 Db 62 CAGGCAAGCTTGGTCTAGCTACATCGACCTGACTGAGTGTCCCTACATGTGGCCCTAC 121
 QY 501 CysSerGlnProTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn 520
 Db 122 TGCTCCCAAGCTTACTACTATGGAGATGCCACATCGTTAATGTACACCTCTCAAT 181
 QY 521 GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln 540
 Db 182 GGCATGGGGCTCAGCAGAGATGTGGATAGCTGAGTGGCGACCTATTATTACACAG 241
 QY 541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGly 560
 Db 242 AATGAGACAAACATTTGAAGTGGCTTCTCTACTCCTCAGTGTGTGGCCCTGTGAGGT 301
 QY 561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln 580
 Db 302 TACCTTGCCATCTCCATTTCTGTGACCAAGAGGACAGTCTCTCGGAGGCATCGCTCAG 361
 QY 581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu 600
 Db 362 GCCACATCATGATCAGTGGCGTCCCAGCAGACAGAGTGTACACAGTGTGGCGAG 421
 QY 601 GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProArgSer 620
 Db 422 CACACTCCACCGTGAAGCTGCCATCAAGTGAAGATCATTCACACCTCTCTCGGAGC 481
 QY 621 LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArg 640
 Db 482 AAGAGAGTCTCTGGGACCATACCAACCTCCGCTACCCACCTGGTCTCTCCCGAG 541
 QY 641 AspaanLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisAsn 660
 Db 542 GACACTTGGGATGAGATGATGACCTTTAGCTGGATGGCAGCAGCTCCACACCAAC 601
 QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
 Db 602 TTCAGGACATGTACCAAGCATCTGCGCAGCATGGGCTACTCTGTGGAGTGTCTGGCGCC 661
 QY 681 -ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluG 700
 Db 662 CCCATTACATGTTTGGACGCCACACAGTATGGCACTTTGCTGCTGGTGGACAGTGAGGA 721
 QY 700 u 700
 Db 722 A 722

RESULT 25

LOCUS BI091766 880 bp mRNA linear EST 20-JUN-2001
 DEFINITION 602858816F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:5000283 5',
 mRNA sequence.
 ACCESSION BI091766
 VERSION BI091766.1 GI:14510096
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 880)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1031 row: o column: 04
High quality sequence stop: 726.

FEATURES

source

1..880
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5000283"
/cell_line="NIH-MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

BASE COUNT 214 a 236 c 239 g 191 t
ORIGIN

Alignment Scores:

Pred. No.: 2,81e-115 Length: 880
Score: 1248.00 Matches: 258
Percent Similarity: 89.86% Conservative: 8
Best Local Similarity: 87.16% Mismatches: 22
Query Match: 22.22% Indels: 10
DB: 13 Gaps: 3

US-09-830-837-6 (1-1052) x BI091766 (1-880)

QY 655 AsphHisIleHisThrAsnPhaArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPhe 674
|||||
DB 3 GATCATCCACACCAATTCAGGGATATGTACAGCACTGAGAGCATGGGCTACTTT 62
|||||
QY 675 ValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeu 694
|||||
DB 63 GTAGAGGTCCTCGGGGCGCCCTTCAGGTGTTTGTATGCCAGTCAGTATGCCACTTTGCTG 122
|||||
QY 695 MetValAspSerGluGluTyrPheProGluGluIleAlaLysLeuArgAspVal 714
|||||
DB 123 ATGGTGACAGTGAGGAGGAGTACTTCCCTCGAGAGATGCCAAGCTCGGAGGGACGTG 182
|||||
QY 715 AspAsnGlyLeuSerLeuValIlePheSerAspTyrTrpAsnThrSerValMetArgLys 734
|||||
DB 183 GACACGGCTCTCGCTCGTCATCTCAGTGACTGTGTACACACTTCTGTATGAGAAA 242
|||||
QY 735 ValLysPheTyrAspGluAsnThrArgGlnTyrTrpMetProAspThrGlyGlyAlaAsn 754
|||||
DB 243 GTGAAGTTTATGATGAAAAACACAGGCAGTGTGGATCCCGGATACCCGAGGAGCTAAC 302
|||||
QY 755 IleProAlaLeuAsnGluLeuSerValTrpAsnMetClyPheSerAspGlyLeuTyr 774
|||||
DB 303 ATCCAGCTCTGAATGAGCTGCTGCTGTGTGGAACATGGGGTTCAGCGATGGCTGTAT 362
|||||
QY 775 GluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAla 794
|||||
DB 363 GAAGGGGAGTTTACCTCGCCACCATGACATGATATATGCGTCAGGGTCAGCATCGCG 422
|||||
QY 795 LysPheProGluAspGlyValValIleThrGlnThrPheLysAspGlnGlyLeu-GluVa 814
|||||

Db

423 AAGTTTCCAGAGATGGCGCTGTATAAACACAGACTTTCAGGACCAAGGATTGGCAGGT 482
QY 814 lLeuLysGlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyrGlnIlePr 834
|||||
DB 483 TTTAAAGCAGAAACAGCAGTGTGTTGAAACGTCCTCATTTTGGGACTTTATCAGATTCC 542
|||||
QY 834 oAlaGluGlyGlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHi 854
|||||
DB 543 AGCTGAGGGTGGAGCGCGGATTCTACTGTATGGGACTCCCAATTGCTTGGATGACAGTCA 602
|||||
QY 854 sArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyVa 874
|||||
DB 603 CCACAGAGAGGACTGCTT-TGGCTTCTGGATGCCCTCTCCAGTACACATCGTATGGGT 661
|||||
QY 874 lThrProProSerLeuSerHisSerGlyAsnArgGlnArgProProSerGlyAlaGlySe 894
|||||
DB 662 GACACCGCTTAGCTCAGTCAGTCTGGGAACCG-CAGCGCTCCCTAGTGAAGCAGGCTC 720
|||||
QY 894 rValThrPro---GluArgMetGluGlyAsnHisLeu---HisArgTyrSerLysValle 912
|||||
DB 721 AGTCAGCTCCAGAGAGGAAATCGGAAGGAAACGATCTTCCATCGGACTCCAAGTTCTCT 780
|||||
QY 912 u-GluAlaHisLeuGly---AspProLysProArg-ProLeuProAla-CysProArg-L 930
|:::|
DB 781 CGCAGGCCCCAATTCGCCCAAGAACCCACAACTCGCGCTCTAGCAACCCCTGCCACGCC 840
|||||
QY 930 euSerTrpAlaLysProGlnProLeuAsnGluThrAla 942
|||
DB 841 TGCCATGGCGCCAGCACCAGCCCTCTACACCCACACCGCC 878
|||||

Search completed: June 2, 2003, 03:52:21

Job time : 3431 secs

THIS PAGE BLANK (USPTO)